A MIXED METHODS STUDY OF TEACHER EVALUATION REFORMS AND MICROPOLITICS IN ILLINOIS

BY

DAVID LEE CONRAD

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Education in Educational Organization and Leadership with a concentration in Educational Administration and Leadership in the Graduate College of the University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Donald G. Hackmann, Chair and Director of Research
Assistant Professor Rachel Roegman
Associate Professor James Shriner
Dr. Linda Sloat
Abstract

National studies have shown that most teachers receive summative evaluation ratings of “satisfactory” or “excellent,” but more are underperforming than evaluation data indicate (Kraft & Gilmour, 2017). Illinois enacted education reforms known as Senate Bill 7 (SB7) and the Performance Evaluation Reform Act (PERA) that required the inclusion of student growth as a significant factor in evaluation processes, and policy advocates called for rigorous teacher evaluations to improve or remove underperforming educators from the classroom (Regenstein, 2011). Since the reforms have been enacted, Illinois policymakers have minimal information to determine whether these reforms have adequately addressed concerns about educator underperformance.

The purpose of this study was to examine the phenomenon of teacher evaluation, focusing on how micropolitics have influenced the implementation of teacher evaluation reforms in Illinois. An explanatory sequential mixed methods design with the follow-up explanation variant was selected to collect data in two phases. The study examined two research questions: (a) to what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools, and (b) how have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms. The study used a conceptual framework based on education policy implementation theory (Honig, 2006) and micropolitics of personnel evaluation (Bridges & Groves, 1999).

In the quantitative phase, 89 superintendents responded to a questionnaire requesting data from 2006-2007 through 2016-2017 on remediation plans, Professional Development Plans, and dismissals in their districts. Findings revealed trends showing small increases in the use of
improvement levers and teacher dismissals following implementation of teacher evaluation reforms, but the number of underperforming educators identified was low compared to estimates of underperformance by evaluators and the literature. In the qualitative phase, 20 principals were interviewed about the influence of micropolitics on their implementation of evaluation reforms. The principals reported that joint committees in their districts created procedures for student growth measures and summative ratings that were favorable to educators, which ultimately increased the teachers’ overall summative evaluation ratings. Second, strategic decisions by evaluators included deferral of low summative ratings due to pending retirements, avoidance or discomfort to hold difficult conversations regarding teacher underperformance, and the increased workload and paperwork involved with the teacher evaluation process and development and monitoring of improvement plans. Finally, principals reported that, although teachers and unions advocated for their interests in designing the procedures, they believed all parties shared a mutual interest in having quality teachers in the classroom. Several recommendations for policy development, professional practice, and further study are presented.
Acknowledgments

I would like to thank my family and my fiancée, Diane Stedman, for their understanding and patience. Being a full-time school leader is time consuming on its own, but adding the commitment of doctoral study was an added stressor my family patiently supported.

I would like to thank my employer and my staff for their support of my work. Our positive and empowering culture has developed leadership “outside the office” with a team effort that shares responsibility for our school. During my doctoral study, my assistant principals were Randall Fortin, John Snipes, and Nathan Short. These leaders each stepped-up to help me when I needed the time for travel, courses, conference presentations, interviews, writing, and the myriad expectations of doctoral study.

My colleagues Dr. Scott Woods and Dr. Sandi Estep were helpful in providing review and validation for the study. Twenty school principals and 89 superintendents helped provide the data for the study. I thank these education colleagues for their time and assistance.

I would like to thank my committee members—Dr. Donald Hackmann, Dr. Rachel Roegman, Dr. James Shriner, and Dr. Linda Sloat—for their time and input in crafting this project. Their guidance is greatly appreciated. In addition, Dr. Jennifer Greene and Dr. S. Kern Alexander provided preliminary feedback for which I am grateful. I especially want to thank my advisor, Dr. Donald Hackmann, who has been a tremendous teacher and mentor to me. He has influenced my research agenda and future career path. I am grateful for his kindness and honesty.
Table of Contents

Chapter 1 Introduction .................................................................................................................. 1

Chapter 2 Review of the Literature ............................................................................................... 16

Chapter 3 Methodology .................................................................................................................. 62

Chapter 4 Results .......................................................................................................................... 76

Chapter 5 Discussion, Implications, and Recommendations ......................................................... 142

References ....................................................................................................................................... 171

Appendix A Data Collection Matrix ............................................................................................... 194

Appendix B Visual Model of Study ................................................................................................ 195

Appendix C Advanced Survey Request ......................................................................................... 196

Appendix D Quantitative Phase Aggregate Data Tables ................................................................. 197

Appendix E Superintendent Questionnaire ..................................................................................... 198

Appendix F Email Confirming Study Eligibility and Soliciting Participation ............................... 205

Appendix G Screening Interview via Telephone .............................................................................. 206

Appendix H Informed Consent Documents .................................................................................... 208

Appendix I Semi-Structured Interview Protocol ........................................................................... 210

Appendix J Institutional Review Board Approval .......................................................................... 212

Appendix K Interview Log ............................................................................................................... 214
Chapter 1

Introduction

Recent public education reforms have been driven by political pressures to increase teacher accountability (Ravitch, 2016). To satisfy this demand, reform efforts have focused on teacher quality, measuring the influence teachers have on student achievement, and the methods used to evaluate teacher performance (Kraft & Gilmour, 2017; Ravitch, 2013). Meanwhile, a growing commitment to leadership for learning has shaped evolving roles and responsibilities for school leaders—as leaders who implement teacher evaluation reforms to improve teaching and learning (Copland & Knapp, 2006; Darling-Hammond, 2013). As their role expectations have become more complex, school leaders have been challenged to balance the numerous and competing responsibilities of leadership, evaluation, management, and supervision (Hazi & Rucinski, 2009) as well as the sheer workload of the position (Grubb & Flessa, 2006). As a result, some school administrators view teacher supervision and evaluation as a bureaucratic duty—something to check off their lengthy list of administrative tasks (Holland, 2004). In addition, the expansion of responsibilities puts increased pressures on leaders to be “super-principals” who must have general expertise in all areas of school leadership to carryout myriad responsibilities (Grubb & Flessa, 2006).

The work to improve teaching and learning practices is time consuming and challenging, and rather than address the performance issues of teachers, some school leaders choose to ignore or minimize the importance of this duty (Dandoy, 2012; Jacob, 2011). The local context and micropolitics within the school may reduce the willingness of school leaders to address teacher underperformance (Blase & Blase, 2002). Numerous political actors influence decisions and implementation of teacher evaluation systems within the school, and rather than to harm
relationships with these actors, school leaders may choose to evaluate teachers with leniency (Kimball & Milanowski, 2009; Reuland, 2012).

To address the economic downtown that began in 2008, the U.S. Congress passed the American Recovery and Reinvestment Act (ARRA) of 2009. ARRA included $5 billion in new funding for competitive Race to the Top (RttT) education grants. States could apply for a RttT grant if they agreed to adopt various policy reforms, including adoption of the Common Core State Standards, inclusion of student growth measures in summative evaluation ratings, expansion of charter schools, and improvement efforts in the lowest-performing schools. Preparing for the RttT applications deadlines, states implemented educational reforms at a much higher rate to satisfy the preconditions under the RttT competition (Howell, 2015; Ravitch, 2013).

In Illinois, the Illinois General Assembly enacted numerous statutory reforms to meet the RttT criteria. Under the Performance Evaluation Reform Act (PERA) enacted in January 2010, statutes were changed to permit additional charter schools, create an electronic longitudinal data system to track student achievement data from pre-K through college, and add student growth measures in teacher and principal evaluations (Illinois State Board of Education [ISBE], 2015). This legislation and subsequent modifications to administrative rules changed the required summative evaluation categories from three categories (excellent, satisfactory, unsatisfactory) to four (excellent, proficient, needs improvement, unsatisfactory), added a required Professional Development Plan (PDP) for any tenured teacher earning a “needs improvement” rating, and mandated a minimum number of formal and informal observations for probationary and tenured teachers (ISBE, 2015). Additional changes included inclusion of student growth measures factored into summative evaluation ratings—to be implemented in all districts by 2016-2017—
and annual summative evaluations for any tenured teacher rated as “needs improvement” or “unsatisfactory,” instead of the 2-year evaluation cycle. A subsequent bill, Senate Bill 7 (SB7), was enacted in June 2011 connecting summative evaluation ratings to layoffs and streamlining rules for local boards of education to dismiss teachers (ISBE, 2015). Upon passage, lawmakers and education advocacy groups hailed these laws as significant reforms that would improve the teaching profession. Some advocates hoped a wider distribution of summative ratings would result, with fewer teachers marked as “excellent” and more teachers rated as “needs improvement” or “unsatisfactory,” and ineffective teachers removed from the classroom (Regenstein, 2011). Historically in Illinois, a minimal number of underperforming Illinois teachers had received “unsatisfactory” summative ratings, and even fewer had been remediated or released for inadequate performance (Reeder, 2005a). Reinforcing this claim in their RttT application (State of Illinois, 2010), Illinois cited data from The Widget Effect (Weisberg, Sexton, Mulhern, & Keeling, 2009) that found among teachers in the three largest Illinois school districts, 92.6% were rated “superior” or “excellent,” 7% were rated “satisfactory,” and only 0.4% were rated “unsatisfactory.”

**Statement of the Problem**

National studies have shown that most teachers receive “satisfactory” or “excellent” summative evaluation ratings, but survey research of school administrators suggest more teachers are underperforming than evaluation data indicate (Kraft & Gilmour, 2017; Weisberg et al., 2008). When Illinois enacted SB7 and PERA reforms, policy advocates called for rigorous teacher evaluations—to improve or remove underperforming teachers from the classroom (Regenstein, 2011). However, a problem exists: Since the reforms were passed, Illinois policymakers have minimal contemporary or longitudinal data to determine whether these
reforms have addressed issues of teacher underperformance. First, comprehensive and complete data on teachers’ annual evaluation ratings, teacher dismissals, and remediation in Illinois are not reported to ISBE by local school districts. Data on teacher underperformance for the past four decades in Illinois is incomplete. Thus, researchers and journalists have utilized inaccurate data on dismissal hearings to estimate teacher underperformance levels (see Henry, 2010; Jones, 1985; London, 1998; Reeder, 2005a; Seltzer, 1992). Second, dismissal data underestimated teacher underperformance numbers, as ineffective teachers resign their positions prior to a state hearing officer—the result of negotiated settlements and/or resignations prior to dismissal (Bridges, 1992; Whitaker, 1999). Finally, the most recent year that dismissal data were collected and reported by researchers was 2008, prior to the enactment of PERA (Henry, 2010). PERA legislation attempted to address this gap by requiring school districts to submit annual comprehensive teacher evaluation data to ISBE, including teacher and principal evaluation ratings from all Illinois public school districts (ISBE, 2015). However, ISBE has been slow to implement this component of the law. The first data were collected during the 2015-2016 school year, but only from the 20% of school districts identified as the lowest achieving and/or districts that voluntarily participated in the Illinois RttT initiative (ISBE, 2016). The first public release of statewide data for all schools occurred in November 2017, reporting on the 2016-2017 school year (ISBE, 2017d).

Although state government bodies can pass laws and establish regulations, the implementation of educational policy still occurs in local schools and districts. Policymakers need accurate data on the effects of policy reforms, but they also need to understand local factors that affect the interpretation and implementation of policy. Local contexts can influence teacher evaluation processes, including the influence of politics within the local school and school
district (Malen, 2006). Therefore, researchers have called for the study of micropolitics—the “formal and informal transactions that shape policy implementation” (Kirst & Wirt, 2009, p. 203)—and its influence on summative teacher evaluations. These micropolitical factors influence how evaluators interact with local personnel and make decisions in the teacher evaluation process (Blase & Blase, 2002; Bridges & Groves, 1999).

To generate an accurate portrait of Illinois teacher evaluation reform, a mixed methods study was appropriate. This study addressed the problems of incomplete data on teacher underperformance and the influence of micropolitics on policy implementation. In the quantitative phase, personnel data were collected on PDPs, remediation plans, and teacher dismissals from all Illinois public school districts to determine if the reforms influenced the identification of underperforming teachers. In the qualitative phase, evaluators were interviewed to determine whether micropolitics affected the identification of underperforming teachers.

**Purpose of the Study**

The purpose of this study was to examine the phenomenon of teacher evaluation, focusing on how micropolitics have influenced the implementation of teacher evaluation reforms in Illinois. An explanatory sequential mixed methods design with the follow-up explanation variant was used. This involved the initial collection and analysis of quantitative data and then expanding upon these findings through in-depth qualitative data from participants who were positioned to assist in illuminating the results (Creswell & Plano Clark, 2011; Ivankova, Creswell, & Stick, 2006). In the quantitative phase, personnel data were collected and analyzed from all Illinois public school districts to assess whether evaluation policy reform (independent variable) affected the identification of underperforming teachers (dependent variables). The qualitative phase was conducted to unpack and provide depth to the quantitative results.
Illinois public school principals were interviewed to explore how micropolitics of personnel evaluation influenced the identification of underperforming teachers. Quantitative data provided a broad overview of the identification of underperforming teachers statewide, while the qualitative data provided specific details on micropolitical factors that influenced principals when identifying underperforming teachers in their local schools.

**Research Questions**

The following research questions guided this study:

**Research question 1:** To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools?

**Research question 2:** How have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms?

**Conceptual Framework**

Two conceptual frameworks informed this study: micropolitics of personnel evaluation (Bridges & Groves, 1999) and education policy implementation theory (Honig, 2006). Education policies are created away from the local site, in state or federal capitals, but the failure or success of policies occurs at the local school level where most policies are implemented (Kirst & Wirt, 2009). Education policy implementation theory describes the interaction of policies, policy actors, and policy implementation sites to understand how policies are implemented (Honig, 2006). Policy change requires modifications to the local school culture, and policymakers often ignore the political realities of local implementation (Hess, 2013). The micropolitics of personnel evaluation framework describes how local policy actors influence the evaluation process at the
local school level (Bridges & Groves, 1999). The struggle is for power and control over “adult issues that supersede anything that goes on in the classroom” (Owen, 2006, p. 4).

**Methodology**

This study included both quantitative and qualitative methods. In the quantitative phase, questionnaires were emailed to 859 public school district superintendent in Illinois to gather data on the numbers of PDPs, remediation plans, and teacher dismissals between the 2006-2007 and 2016-2017 school years. In addition, data were obtained from ISBE on the numbers of educators employed and student enrollments between the 2006-2007 and 2016-2017 school years. Additional demographic information, including the percentage of low income students, the percentage of students belonging to a particular racial/ethnic group, and contact information for district and school leaders for the 2016-2017 school year, was obtained.

Information from the quantitative first phase was further explored in a second qualitative phase. In-depth semi-structured interviews were conducted with 20 practicing evaluators from Illinois public schools who had experience before and after the policy reforms and who had identified an underperforming teacher. These interviews addressed micropolitical factors that may have influenced the principals while carrying out the evaluation reform policies in their local schools. The findings of the quantitative and qualitative phases were integrated during the interpretation phase of the mixed methods study (Creswell & Plano Clark, 2011).

**Researcher’s Statement**

For 11 years, I have served as a middle school principal in an Illinois public school district, during which time I have evaluated teachers annually. As a doctoral student/researcher and practitioner, I am interested in the issues surrounding supervision and evaluation and the identification and support of underperforming teachers. In my leadership practice, I have
conducted countless observation cycles; written numerous summative evaluation reports; made recommendations for employment, re-employment, and dismissal; and led and/or participated in the writing and implementation of PDPs and remediation plans for educators. When confronted with these high-stakes decisions, I have been challenged by the politics that influenced my decision making.

Without the support of my superintendent and local board of education, these plans would not have been written. They establish and reinforce the parameters for an organizational culture demanding high-quality teachers for every student. However, I have experienced micropolitical pressures from local actors—including the teachers’ union, community and family members within the district, board of education members, and staff relationships—who interacted to influence my efforts to remediate underperforming teachers. Some actors were supportive, while others were opposed. In some cases, I was slower and more cautious in addressing underperformance, fearing disruptions to the school climate and fractured relationships between staff and myself. Due to my ongoing experiences, I have often wondered whether other school leaders experience similar challenges and, if so, how local politics influences their decisions.

Limitations

A perfect study is not possible. Therefore, the researcher must delineate the limitations associated with the study (Mertens, 2009). The teacher underperformance data used in this study was limited to ratings of all Illinois public school educators who are subject to Illinois teacher evaluation laws and are in non-administrative appointments. These personnel may include general education and special education teachers, counselors, social workers, certificated nurses,
psychologists, library/media specialists, and instructional coaches, among other certificated positions.

The accuracy of the quantitative findings depended upon the accurate reporting of Illinois school districts subject to the FOIA request. Errors could include data entry, inaccurate calculations, and misunderstanding of the data collection criteria and procedures. In addition, resignations may skew the accuracy of the data. For example, teachers might resign their position due to a negotiated settlement, upon obtaining other employment, to avoid discipline, or to avoid initiation of a PDP or remediation plan.

One limitation was the quantitative data did not reflect the identification of all non-tenured teachers in Illinois. First, under Illinois statutes, Illinois school boards may remove a non-tenured teacher without the requirement of initiating a PDP or remediation plan (ISBE, 2015); they may simply non-renew the teacher’s annual contract. Second, tenured and non-tenured teachers may resign to avoid dismissal and will not be reflected in the quantitative dismissal, PDP, or remediation data. Finally, the voluntary questionnaire only represented only a sampling of Illinois public school districts; therefore, results cannot be generalized as representative of the entire state.

The accuracy of the qualitative findings was dependent upon the willingness of participants to participate and provide candid and honest responses to the interview questions. First, participants self-selected or volunteered to participate in the interviews, which may have created self-selection bias influencing the findings (Robinson, 2013). Second, gaining access to interview participants was difficult if sensitive information about school politics could have been perceived as damaging for a potential study participant (Flessa, 2009). Third, a participant may have provided a response that he/she deemed as socially desirable, rather than one that was
representative of one’s true opinions. Fourth, the interview participants included only a small sample of the total number of evaluators in Illinois; also, the experiences of participants in the qualitative phase may not have been representative of the school districts whose data were reported in the quantitative phase. Although I hoped to obtain participants employed by the respondent districts from the quantitative phase, the criteria for inclusion—including 7 years of evaluation experience and experience implementing an improvement lever—resulted in a small pool of potential participants. In addition, member checks were conducted post-analysis. Therefore, the experiences of participants cannot be generalized as representative of all school districts or evaluators in Illinois.

**Delimitations**

This study was delimited to the study of the PERA and SB7 teacher evaluation reforms in Illinois. Teacher evaluation policies and practices vary across states, and the findings from this study may not apply to other states whose laws and implementation processes may differ. For example, Illinois statutes reflect more comprehensive requirements for school districts to bargain or consult with their local teachers’ unions on various aspects of the evaluation process than may be the case in many other states. The results of this study may not apply to states in which collective bargaining does not influence local evaluation policy.

The population in the qualitative phase included 859 public elementary, high school, and unit districts in Illinois. The study excluded other public school entities, including the Illinois Department of Corrections statewide school district, 13 state-funded schools including university-affiliated schools and state-run facilities for special needs students, 110 special education districts, 77 vocational/technical cooperatives, and 38 Intermediate Service Centers and Regional Offices of Education that supervise and operate state-funded schools and regional
programs. Teacher evaluation processes in Illinois private and parochial schools are not regulated by Illinois law and therefore were excluded from the study.

The participants in the qualitative phase were delimited to principals in Illinois public schools who have undertaken specialized training and assessments and engaged in the practice of evaluation. To become qualified evaluators, Illinois evaluation reforms required all potential evaluators to complete an online training course and pass interrater reliability exams (ISBE, 2015). However, evaluators could include those working in positions besides principals and assistant principals—department chairpersons, teacher leaders, content area leads, and others so qualified and assigned to evaluate educators. Under Illinois law, qualified evaluators are not required to hold an administrative certification in order to evaluate teachers.

**Significance of the Study**

This study is significant because of the need to evaluate recent Illinois education policy reforms and the effects of their implementation. Using the frameworks of micropolitics in education and education policy implementation theory, the study informed Illinois teacher evaluation policy and practice by identifying the supports and barriers that influence evaluators when identifying underperforming teachers. In addition, Illinois makes an interesting case for studying teacher evaluation and policy reforms. The negotiation process for developing PERA and SB7 reform legislation was heralded as a new model for collaboration between advocacy groups, legislators, policymakers, and teachers’ unions (Regenstein, 2011). Thus, this study provided a mechanism to determine the extent to which this collaboration yielded the intended implementation outcomes. Finally, the findings can be useful, to help shape future policy and provide feedback to legislators and stakeholders on the effects of legislation.
**Definition of Terms**

**Advocacy group.** A non-profit organization—including corporate reformers (Ravitch, 2013), philanthropies, sociopolitical organizations, think tanks, and labor unions—that engages in political influence in education policy by using strategies such as publishing research reports, funding pilot projects, and/or engaging in direct political influence in public policymaking (Scott, Lubienski, & DeBray-Pelot, 2009).

**Certificated staff.** In Illinois public schools, certificated staff includes teachers, administrators, school psychologists, social workers, school counselors, and school librarians holding specified teaching certifications through ISBE (2015). For the purposes of this study, the term “teacher” and “educator” are synonymous with “certificated staff”—but excluding educators employed under administrative contracts.

**Illinois State Board of Education.** The Illinois State Board Education (ISBE) is the state agency responsible for creating administrative regulations for K-12 education based upon Illinois statutes. This agency oversees public school districts in Illinois, collecting annual school data, dispersing education funding, and overseeing regulatory compliance.

**Performance Evaluation Reform Act.** Public Act 96-861, also entitled the Performance Evaluation Reform Act (PERA), is an Illinois statute passed in January 2010, which contained a variety of education reforms focusing on the evaluation of educators and principals, including student growth measures for teachers and principals (ISBE, 2015).

**Political actors.** Political actors are persons or groups “with their resources seeking to accomplish distinctive goals within a context where conflict regularly prevails” (Kirst & Wirt, 2009, p. 204). They may include philanthropic groups, advocacy organizations, educational interest-based groups such as labor unions or parent/teacher groups, administrators, professional
organizations, corporations and businesses, bureaucrats, legislators, teachers, students and parents (Kirst & Wirt, 2009; Marsh & Wohlstetter, 2013; Scott et al., 2009).

**Policy implementation.** Educational policy implementation describes the interactions of policies, people, and places and how they shape the implementation result (Honig, 2006).

**Professional Development Plan.** Under the Illinois PERA, a tenured teacher who receives a “needs improvement” summative rating must undertake a Professional Development Plan (PDP) for the following school year that contains improvement activities. Regulations have defined this plan to include frequent observations, classroom supports, and specific activities to improve performance (ISBE, 2015).

**Qualified evaluator.** Any Illinois school educator who is approved to evaluate public school educators. Starting in the 2012-2013 school year, qualified evaluators were required to pass a comprehensive training program that contains information on the *Danielson Framework for Teaching* (Danielson, 2013), video observations, and content about Illinois teacher evaluation rules and procedures. Evaluators are not limited to principals, assistant principals, and other job titles with administrative contracts, but may include department chairpersons, teacher leaders, instructional coaches, and other non-administrative educators who have met the qualifications to evaluate educators (ISBE, 2015). Qualified evaluators may include any school positions that are responsible for evaluating educators—including principals, assistant principals, department chairpersons, directors, teachers, and others given the authority to engage in teacher evaluation in their districts. As of April 2018, 16,083 educators had passed the prequalification tests to conduct teacher evaluations with student growth measures in Illinois, but the number who actually conduct evaluations is much less (Growth Through Learning, 2018).
**Race to the Top.** Under the Obama administration, the federal government awarded $5 billion in the Race to the Top (RtT) competitive grant program to states who applied and agreed to implement various educational initiatives addressing data management, college readiness, teacher effectiveness, and low-performing schools (Howell, 2015; Superfine, Smylie, Cummings, & Tozer, 2012). Illinois was awarded $43 million in the third round of the RtT competition (Howell, 2015; State of Illinois, 2010).

**Remediation plan.** Under the Illinois Performance Evaluation Reform Act (PERA), tenured teachers who receive an “unsatisfactory” rating must complete a remediation plan. Illinois statutes have defined this plan to include frequent evaluation, classroom support, a consulting teacher, and specific activities to improve performance. Boards of education may initiate a state mandated process to dismiss a teacher who remains “unsatisfactory” after completing the remediation plan (ISBE, 2015).

**Senate Bill 7 (SB7).** In Illinois, Public Act 97-0008 was enacted in June 2011. This legislation was connected to PERA, extending summative evaluation ratings to tenure and layoff decisions, as well as revising negotiation procedures between boards of education and their local teachers’ unions (ISBE, 2015).

**Underperforming educator.** For the purposes of this study, an educator who receives a “needs improvement” or “unsatisfactory” rating shall be considered underperforming. In Illinois, these ratings mandate a PDP or a remediation plan (ISBE, 2015).

**Summary**

This dissertation includes five chapters. Chapter 1 provided an overview of the study, including an introduction to the problem, purpose, research questions, methodology, and significance. Chapter 2 contains a review of the literature about educator evaluation ratings and
the factors that influence them. Chapter 3 describes the research questions for the study, a
description of the methodology, the population, sample selection, data collection methodology,
and data analysis procedures that will be utilized. Chapter 4 provides an overview of the findings
from the study. Chapter 5 focuses on interpretation of the research findings as related to the
conceptual framework and literature review. Recommendations for policy, practice, and future
research are discussed in detail.
Chapter 2

Review of the Literature

Over the past 30 years, state and federal policymakers have implemented myriad reforms intended to improve public education. Faced with political pressures and federal incentives to change, state policymakers enacted legislation to improve instruction through teacher evaluation reforms to create the conditions for school improvement (Darling-Hammond, 2014; Ravitch, 2016). In 1983, *A Nation at Risk* called for higher standards for teacher preparation, accountability for teachers, and teacher pay structures tied to quality and performance (National Commission on Excellence in Education, 1983). Described as “the all-time blockbuster of education reports” (Ravitch, 2013, p. 27), this task force decried that America’s schools were failing and in need of significant reforms. In response to public reaction to the report, some states initiated new systems of teacher evaluation, including enhanced evaluator training, new teaching standards, and early efforts at performance evaluation reforms (Hazi & Rucinski, 2009).

Influenced by the reactions to *A Nation at Risk*, Illinois enacted the 1986 School Reform Act 24A, changing the process for teacher remediation and dismissal (London, 1998; McDonald, 1992). Several years later, the federal No Child Left Behind Act of 2001 was passed with the focused purpose of raising student test scores in reading and mathematics (Ravitch, 2013). This legislation implemented higher standards for new and current teachers, defining new standards for “highly-qualified” teachers in every classroom. The law also implemented mandatory student testing for accountability purposes, annual performance goals for schools, and the requirement that 100% of students meet proficiency standards by 2014 (No Child Left Behind Act of 2001 [NCLB], 2002). In 2009, the RttT grant program (ARRA, 2009) provided financial incentives for states to implement new plans for college and workplace readiness, data systems to track student
growth, methods to address teacher effectiveness, and reforms for the lowest performing schools. RttT lured many states to implement reforms to satisfy the requirements for their grant applications, including teacher evaluation reforms. These changes included connecting teacher evaluation to student growth, requiring new professional development for teachers and principals, linking compensation to teacher retention/promotion, the intent to produce rigorous rules for granting tenure and certification, and rules for removing underperforming teachers from the workplace (Howell, 2015; Lavigne & Good, 2014). When Illinois enacted reforms to meet the RttT grant conditions, some advocacy groups hoped for outcomes that included the improvement or removal of underperforming teachers (Regenstein, 2011).

This chapter reviews the literature on teacher evaluation and the role it plays on improving or removing underperforming teachers. First, research on leadership for learning describes the current focus of school leaders on teaching and learning. Next, the historical development and purposes of teacher evaluation are explored, followed by a review of teacher quality. The distribution of teacher evaluation ratings shows national data and trends on summative evaluation ratings, followed by the history of teacher evaluation processes in Illinois. To provide explanations for these outcomes, the literature on factors that influence evaluators when identifying underperforming teachers is explored to learn why evaluators may be influenced to issue performance ratings that do not align with actual teacher performance. Finally, to frame this study, the conceptual frameworks of micropolitics of personnel evaluation (Bridges & Groves, 1999) and education policy implementation (Honig, 2006) are presented.

**Leadership for Learning**

Throughout the past two decades, the practice of school leadership has shifted from responsibilities normally considered managerial toward models emphasizing student learning.
This shift has challenged school leaders to balance the often-conflicting roles of management tasks and instructional leadership (Grubb & Flessa, 2006; Hallinger & Murphy, 2013; Portin & Knapp, 2014). Meanwhile, research on school leader practices began to show correlations between effective leadership and improving student learning, recognizing that certain leadership behaviors were more likely to be observed in higher performing schools (Copland & Knapp, 2006; Murphy, Hallinger, & Heck, 2013). A review of literature found leadership to be the second most important in-school factor influencing student achievement, second only to the classroom teacher (Leithwood, Lewis, Anderson, & Wahlstrom, 2004). Research estimated highly effective school leaders can add two to seven months of achievement in mathematics when effective leaders support the conditions for learning. This effect occurred as quickly as one academic year when effective leaders were compared to ineffective leaders (Branch, Hanushek, & Rivkin, 2013).

Leaders create the conditions for learning in which “individual variables combine to reach critical mass” (Wallace Foundation, 2013, p. 5). In response to the emerging research on these variables, Copland and Knapp (2006) created a leadership for learning framework. Murphy, Elliot, Goldring, and Porter (2007) subsequently developed a taxonomy of dimensions and functions to define leadership for learning, including creating a vision for learning, instructional program, curricular program, assessment program, communities of learning, resource allocation and use, organizing culture, and social advocacy. A complementary model by Hallinger (2011) described leadership for learning through the four dimensions of values leadership, leadership focus, context for leadership, and sources of leadership. Although multiple terms for these models of this leadership exist—including leadership for learning, instructionally focused leadership, and leadership for school improvement—each model emphasizes the
essential role of the principal in promoting teaching quality and learning (Murphy et al., 2007). The next section will describe the leadership for learning framework through the five conditions leaders establish to support learning: (a) establishing a focus on learning, (b) building professional communities that matter for learning, (c) engaging external environments that matter for learning, (d) acting strategically and sharing leadership, and (e) creating coherence (Copland & Knapp, 2006). These conditions are addressed in this section.

**Establishing a focus on learning.** Leadership for learning calls on school leaders to frame and embrace a common vision for learning focused on the alignment of curriculum, instruction, and assessment to promote student growth (Copland & Knapp, 2006). This vision is crucial, as “the ability to articulate a learning focused vision that is shared by others and to set clear goals creates a base for all other leadership strategies and actions” (Hallinger, 2011, p. 137). The vision must come alive as the leader models and articulates its meaning to everyone in the school community (Copland & Knapp, 2006).

Copland and Knapp (2006) described actions that school leaders can undertake to improve learning focus. Leaders for learning are visible in schools with frequent classroom observations—on a daily or weekly basis—to support teaching and professional development. Leaders use these observations to engage in professional conversations about instructional practices (Portin & Knapp, 2014). Research has found that administrative time spent on evaluations and coaching teachers for improvement is correlated to improved student achievement (Grissom, Loeb, & Master, 2013). Learning improvement is driven by assessment data from students to evaluate the effectiveness of the instructional program. Leaders for learning help teachers make sense of the assessment data through conversations and professional development. Both leaders and teachers develop plans to improve the results by setting goals and
planning strategies for improvement (Murphy et al., 2007). Leaders for learning provide continual staff development experiences for teachers by monitoring performance and targeting assistance based on data that informs areas of need (Hallinger, 2011).

**Building professional communities that matter for learning.** Leaders for learning create a culture where a learning is valued by everyone in the school community. They promote this focus by visibly demonstrating their commitment to learning (Copland & Knapp, 2006). Leaders organize teachers in ways that promote natural collaboration and time for structured group work, including the formation of instructional teams that lead instructional improvement (Portin & Knapp, 2014). Effective leaders utilize the best models and research to provide quality staff development experiences (Murphy et al., 2007). Leaders for learning also hire and retain highly effective teachers whose values fit the culture of the school (Branch et al., 2013).

Leaders must take specific steps toward building trusting relationships with their faculty and staff. This work includes changing the supervisory relationships between teachers and leaders, emphasizing informal classroom visits beyond the formal evaluation process. Portin and Knapp (2014) described several methods leaders utilize to promote this culture. Leaders for learning engage in informal conversations with teachers in hallways and outside of the school day. Leaders also promote an open-door policy to encourage teachers to share ideas freely and engage in professional learning conversations. In addition, leaders demonstrate a positive, optimistic tone that welcomes engagement from all members of the school community.

**Engaging external environments that matter for learning.** Leaders for learning engage the external community to create opportunities for learning. This external community includes parents, community and neighborhood groups, advocacy groups, news media, taxpayers, and local governments. This engagement includes efforts to secure necessary funding
and other resources to improve learning in the school. Leaders also manage the politics of schools through relationships with outside policy actors. These policy actors provide opportunities for both support and resistance, and leaders for learning employ proactive strategies to build relationships with these actors, publicly promoting the vision of their school (Copland & Knapp, 2006; Kirst & Wirt, 2009).

**Acting strategically and sharing leadership.** Leaders for learning strategically distribute leadership to teachers to improve learning. Leaders empower teachers to make effective decisions about the operations and management of the school (Copland & Knapp, 2006; Portin & Knapp, 2014). Teachers who are empowered to select their own professional learning goals may experience increased professional improvement (Sandholtz & Scribner, 2006). In addition, collaboration that distributes power between the teachers and leadership may improve implementation of teacher evaluation policies (Behrstock-Sherratt, Rizzolo, Laine, & Friedman, 2013). Teacher input to design the evaluation process can result in higher implementation fidelity and increased validity of the evaluation results (Goe, Bell, & Little, 2008).

**Creating coherence.** Leaders for learning also promote clarity and coherence in their efforts to improve learning. Within the school, coherence occurs on three levels. First, coherence involves the alignment of methods, activities, and resources to ensure quality instruction. Second, coherence involves the alignment of leadership and a common vision. Third, coherence requires a shared consensus between all members of the school community (Copland & Knapp, 2006).

Coherence is challenging when policies create conflicting demands on schools, especially when state and federal policies intersect with local initiatives. Many schools are challenged by myriad programs that work individually but do not align with the common vision (Superfine et
In their drive for learning improvement, leaders must engage their school community to ensure learning improvement efforts are working congruently towards the common goals:

Improvement is the change with direction, sustained over time, that moves entire systems, raising the average level of quality and performance with at the same time decreasing the variation among units, and engaging people in analysis and understanding of why some actions seem to work and others don’t. (Elmore, 2000, p. 13)

In the next section, the history of teacher supervision and evaluation will be explored, including the development of the purposes of these functions.

**History of Teacher Supervision and Evaluation in the United States**

Current educational literature defines teacher supervision through the lens of instructional supervision—a nonjudgmental process to help teachers improve their instructional skills. Hazi and Rucinski (2009) defined supervision as a professional development process to improve teacher quality. Marzano, Frontier, and Livingston (2011) defined supervision as “the enhancement of teachers pedagogical skills, with the ultimate goal of enhancing student achievement” (p. 2), citing research noting positive correlations between teacher skill and student achievement. Mette et al. (2017) described supervision as “focused on ongoing support, teacher improvement, and teacher professional growth” (p. 710) through formal and informal means.

The process of teacher evaluation has evolved into a legally mandated and regulated personnel function to conduct observations and determine summative evaluation ratings over a defined time period. Although some researchers and practitioners conflate supervision and evaluation as the same activity, they should be considered as separate functions (Hazi & Rucinski, 2009). In general, teacher evaluations included elements of classroom observations, analysis of observation data, and interaction between the teacher and the observer. To conclude the evaluation process, the evaluator makes a summative judgement about the teacher’s performance and issues a final rating or score, potentially affecting future employment status.
(Darling-Hammond, 2013; Mette et al., 2017). Early teacher evaluation was defined and regulated at the local level. However, following the release of *A Nation at Risk* in 1983 (National Commission on Excellence in Education, 1983) and later the passage of the No Child Left Behind Act of 2001, some states developed policies to connect evaluations with teacher accountability (Hazi & Rucinski, 2009). In 2009, through the RTtT grant program, states were lured to reform their teacher evaluation policies to connect student growth scores to summative evaluation ratings, increase the rigor of teacher evaluation instruments, and use evaluation ratings to remove underperforming teachers from the classroom (Howell, 2015; Lavigne & Good, 2014).

Tracy (1995) described seven phases in the historical development of teacher supervision in the United States. The first phase of community accountability occurred between the mid-1600s and the early 1800s. Local committees of government officials or clergy supervised schools, with clergy playing a strong role because of their knowledge of religious curriculum, their own educational attainment, and the cultural norms of the community. These laypersons visited the school to inspect the content being taught, student progress, student discipline, teaching methods, and the physical classroom environment, with inspections occurring annually or monthly (Tracy). Criteria and standards for teacher performance did not exist, resulting in varied quality and types of feedback to teachers. In this phase, however, the primary role was helping the teacher improve teacher performance (Marzano et al., 2011).

The second phase of professionalism occurred during most of the 1800s. School supervision emerged as a specialized role for professionals who were experts in teaching and content (Tracy, 1995). The growth in urban school systems spurred this movement, as schools consolidated into larger districts where specialized roles for content specialists and
administrators could be supported. This trend gradually influenced rural schools, as laypersons and clergy were determined to lack the knowledge and skills to measure teaching effectiveness (Marzano et al., 2011). Teacher institutes and college training were created to provide training to groups of teachers efficiently, but this decreased the individual assistance once provided to individual teachers (Tracy, 1995).

The third phase was the scientific phase, occurring between the early 1900s through the 1920s (Tracy, 1995). The work of Frederick Taylor influenced schools to adopt organizational practices from industry—known as the factory model—to increase productivity. Data could be collected to measure the performance of teachers and schools, including the early use of standards and aptitude tests to measure student learning (Lavigne & Good, 2014; Marzano et al., 2011). In addition, the first teacher ratings scales were developed in 1915 (Lavigne & Good, 2014). Teacher assignments became more specialized into content areas as school enrollment increased, similar to the division of labor on the assembly line (Tracy, 1995). The concept of an organizational structure containing an 8-year elementary school and 4-year high school also was developed (Fine, 1997).

The fourth phase was human relations. Between the 1930s and 1940s, a greater focus on teachers as individuals emerged (Tracy, 1995). School supervision focused on both the professional and personal needs of teachers, including professional development and emotional support. Teachers were afforded greater autonomy and creativity in their classrooms. The roles of supervisors also widened to include curriculum, public relations, school lunches, and other management functions (Marzano et al., 2011). The dominant model for the next century of teacher evaluation was established: A school leader observes a teacher’s lesson using a checklist
or ratings tool, followed by a conference where the evaluator lists the teacher’s areas of strength and weakness (Lavigne & Good, 2014).

The next two phases reemerged from the struggle to balance the emerging knowledge on the elements of effective teaching and the desire to maintain teacher autonomy within their classrooms (Marzano et al., 2011). Between the 1940s and 1960s, the fifth phase experienced the reemergence of the scientific phase (Tracy, 1995). In this era of clinical supervision, new teaching models emerged. The complexity of evaluations increased with prescriptive criteria and procedures to measure the effectiveness of teaching practices (Lavigne & Good, 2014; Marzano et al., 2011). The sixth phase returned to the human relations phase between the mid-1960s and mid-1980s (Tracy, 1995). Evaluators measured teaching performance through the lens of prescribed teaching models, using checklists to ensure the teacher followed the procedures and sequence of the models. These tools gave more direction and structure. This movement was driven by teachers’ unions who pushed for more objectivity—clear and consistent evaluation criteria—in teacher evaluation systems (Kersten & Israel, 1995). The clinical evaluation cycle of pre- and post-conferences emerged during the early 1980s, along with shared responsibility for teacher improvement between the teacher and their supervisor (Hazi & Rucinski, 2009).

The seventh phase of human development started in the mid-1980s. An increased focus on the needs of teachers as adult learners emerged, including concerns over their career stages and personal development (Tracy, 1995). Some research proposed using different indicators of quality to define teachers at different stages of their career (Clayton, 2013). Differentiated evaluation and professional development models emerged to provide targeted supervision based on the experience level and needs of teachers, embracing multiple evaluation models and approaches to fit the needs of the teacher (Marzano et al., 2011). In addition, teacher reflection in
evaluation emerged during this phase (Hazi & Rucinski, 2009). Meanwhile, reports such as *A Nation at Risk* were published to highlight problems in education and call for education reforms (NCEE, 1983). State legislatures—including Illinois—began to pass laws to reform and standardize teacher evaluation practices, including new requirements to train evaluators. The amount of time and effort evaluators devoted to teacher evaluations increased the demands on school leaders (Kersten & Israel, 1995).

Following the second human development phase, an additional teacher evaluation phase has emerged. Since 2009, over two-thirds of states engaged in significant teacher evaluation reforms (Hull, 2013). The role of supervision evolved into a comprehensive teacher evaluation process with alignment to uniform and rigorous teaching standards, a stronger focus on student learning outcomes, and multiple measures of teaching and learning (Donaldson & Papay, 2014; Lavigne & Good, 2015). Charlotte Danielson published her teaching framework in 1996, and it was considered the most detailed model of teaching and teacher evaluation available (Marzano et al., 2011). Danielson’s framework is the most popular model used by schools today (Donaldson & Papay, 2014) and was the teacher evaluation model suggested by Illinois state policy officials (ISBE, 2015). Other models used in Illinois and other states included the core propositions of National Board for Professional Teaching Standards, the Causal Teaching Evaluation Model developed by Robert Marzano, and the Classroom Assessment Scoring System (CLASS; (Donaldson & Papay, 2014; Lavigne & Chamberlain, 2017; Marzano, Carbough, Rutherford, & Toth, 2014).

Many teacher evaluation reforms connected summative evaluation ratings to student scores on student tests (Lavigne & Good, 2014). Some reforms went further to connect teacher salaries to evaluation ratings and student test scores (Goldhaber & Walch, 2011). In 2013, 38
states required summative teacher evaluations to include or consider student achievement scores, while 23 states required student achievement to comprise at least half of the summative evaluation score (Hull, 2013). One early reform connecting student test scores to evaluations was the Tennessee Value Added Assessment System (TVASS), mandated by the 1992 Education Improvement Act in Tennessee (Sanders & Horn, 1994). The TVASS model uses a pre-test and a post-test for each measured course. Using a large database of present and historical information, statisticians developed a scoring model to measure the influence the teacher had on their students’ growth. This reform was attractive to policymakers because it quantified teaching into understandable, objective measures that could be reported to the public and used for accountability purposes (Hanushek, 2009; Mead, Rotherham, & Brown, 2012). Sanders’ early value-added models did not include controls to account for students’ demographic context because TVAAS argued the demographic information for a child’s performance already affected the pre-test scores. A student whose socioeconomic conditions affected her/his assessment scores already was experiencing these conditions at the time of the pre-test and in the longitudinal data collected over time (Ballou, Sanders, & Wright, 2004).

Many studies suggested value-added models were statistically reliable and valid (Ballou et al., 2004; Sanders & Horn, 1998). Others found value-added measures could predict the future performance of both students’ and teachers’ performance (Aaronsen, Barrow, & Sanders, 2007; American Statistical Association [ASA], 2014), predict high school students’ college readiness (Smalskas, 2013), and correlate with teachers’ content knowledge and instructional quality (Hill, Kapitula, & Urnland, 2011). Value-added measures were effective in other applications, including the assignment of students to specific teachers based on instructional and learning styles or content knowledge (McClellan, Donoghue, & Pianta, 2014), to ensure a student was not
assigned to an underperforming teacher for consecutive years (Mead et al., 2012), as the measurement criteria for salary systems connected to student test scores (Goldhaber & Walch, 2011), and for allocating professional development resources to the teachers who needed them most (Ruzek, Hafen, Hamre, & Pianta, 2014).

Other researchers, however, have questioned the utility of value-added measures (Baker, Oluwole, & Green, 2013). Researchers criticized the statistical methodology of TVASS, including the failure to account for prior student learning (Stronge, Ward, & Grant, 2011) and student socioeconomic status (Amrein-Beardsley, 2008; Mead et al., 2012). Sanders’ work initially excluded demographic characteristics from the calculations, but later models incorporated student characteristics such as poverty and race into the calculations (Ballou et al., 2004). Value-added measures showed low reliability, calling into question their use for high-stakes decisions such as teacher dismissal or tenure (ASA, 2014; Staiger & Kane, 2014). Value-added measures may identify the strongest or weakest teachers but lack the sensitivity to distinguish between teachers whose scores lie outside these extremes (Mead et al., 2012). Greater reliability would result from using measurements throughout a teacher’s career to reduce the variability of year-to-year factors (ASA, 2014; Staiger & Kane, 2014). Value-added measures also suffer from low instructional sensitivity, described as the correlation of teaching inputs and student learning outputs (Goe et al., 2008; Polikoff, 2014), partly caused by unclear or divergent indicators of effective teaching (Darling-Hammond, 2013; Kimball & Milanowski, 2009). In the next section, the purposes of the teacher supervision and evaluation will be explored.

**purposes of Teacher Supervision and Evaluation**

The current phase of supervision expands the role and importance of teacher evaluation processes. Besides rating teacher performance, evaluations should provide a comprehensive
system to connect professional development with teacher evaluation (Darling-Hammond, 2013; Goldrick, Zabala, & Burn, 2013). Darling-Hammond (2013) described an overall vision of an ideal evaluation system:

What this country really needs is a conception of teacher evaluation as part of a teaching and learning system that supports continuous improvement, both for individual teachers and for the profession as a whole. Such a system would enhance teacher learning and skill, while at the same time ensuring that teachers who are retained and tenured can effectively support student learning throughout their careers. (p. 5)

In the next section, the purposes of teacher evaluation will be explored using the purposes delineated by Haefele (1993). He called for evaluation reforms to focus on formative and summative aspects of evaluation with the purposes of (a) screening teachers, (b) providing constructive feedback to individual teachers, (c) rewarding outstanding performance, (d) informing professional development, (e) providing legally defensible documentation, (f) terminating incompetent or unproductive personnel, and (g) unifying teachers and administrators in a common effort of educating students. The next section will explore the literature on each purpose.

**Screening teachers.** Screening teachers for rehiring or continued employment is an essential summative function of teacher evaluation (Hanushek, 2009; Staiger & Rockoff, 2010). Leaders for learning apply this understanding to hire effective teachers, provide them the resources and supports to improve instruction, and support their ongoing development (Mason & Schroeder, 2010). Having a quality teacher in the classroom is important due to the influence effective teachers have on student achievement (Guarino, Santibañez, & Daley, 2006; Little & Miller, 2007). The presence of ineffective teachers is a significant reform challenge facing low-achieving schools. Research shows low-achieving schools experience a high teacher turnover rate, making long-term reforms challenging to sustain (Garcia, Slate, & Delgado, 2009; Jacob, 2007). The research connecting these teaching and leadership inputs to student learning helped
advance numerous reform movements around evaluation and improving or removing underperforming teachers (Ravitch, 2013).

A teacher’s academic performance in high school or college may predict teacher quality and higher student achievement (D’Augustino & Powers, 2009). In the United States, only 23% of teachers graduated in the top third of their high school graduating class. In contrast, Singapore limits enrollment in teacher preparation programs to the top 30% of high school graduates. School officials in the United States are hindered in their ability to hire effective teachers when the labor pool includes many teachers who were moderate to low academic performers in high school and college (Auguste, Kihn, & Miller, 2010). Recent pressures from school reforms, decreased public school funding, and career teacher earnings has reduced the numbers of undergraduates who choose to graduate from college with a teaching degree (Sawchuk, 2014).

For early career teachers, Staiger and Rockoff (2010) found quality teachers are best identified after they are hired, collecting evidence throughout their first year of teaching. They argued school leaders should aggressively monitor and support teachers in their early career and only grant tenure to those staff who are excelling as teachers. The importance of these decisions is magnified in states where tenure rights are granted within a short time after initial employment (Weisberg et al., 2009).

Nationwide, teachers in 46 states can earn tenure after 1-5 years of employment, while teachers in four states do not have tenure rights (Thomsen, 2014). Since the passage of SB7 in Illinois, tenure can be granted after 3-4 years, depending on their ratings, but boards of education have an option for granting early tenure after 2 years for teachers were previously tenured in another district and who earned “excellent” ratings for 2 consecutive years. However, most tenure decisions in Illinois are made during a teacher’s fourth year (ISBE, 2015). Nationwide,
few early-career teachers are ever identified as underperforming—and few early-career teachers are denied tenure in their last year of tenure eligibility—questioning the rigor of tenure acquisition (Weisberg et al., 2009).

**Feedback to individual teachers.** School leaders influence student learning and improve learning through feedback to teachers (Copland & Knapp, 2006). Formative feedback is an effective means to improve teaching performance, as one meta-analysis found formative feedback to teachers to have a .90 effect size on student achievement (Hattie, 2009). Feedback helps teachers understand how they are influencing students and student learning, and helps them to learn strategies and methods to improve their performance (Marzano et al., 2011). Danielson (2010) noted, “evaluator-teacher conversations, when conducted around a common understanding of good teaching—and around evidence of that teaching—offer a rich opportunity for professional dialogue and growth” (p. 39).

Supervision and evaluation are more effective when leaders provide honest and specific feedback throughout the supervisory process (Copland & Knapp, 2006). For teachers to improve, they must have accurate information on the effectiveness of their classroom practices and how to improve (Firestone, 2014). Teachers have found specific feedback useful for their improvement, and when such feedback was offered in post-conferences, teachers were more likely to engage in activities to improve their professional practice. These practices included training, professional reading, peer observation, and experimenting with new strategies (Tuylens & Devos, 2011). Research also has shown leaders must build positive, trustful relationships with their staff, in order to be perceived as supportive and genuinely invested in helping them improve their performance (Copland & Knapp, 2006; Fullan, 2014).
Some school cultures are not receptive to honest and challenging discussions about teacher performance (Mead et al., 2012). Teachers may fear summative evaluations because of their potential to create job loss (Conley & Glasman, 2008). Summative evaluation ratings are required by state laws to rate and rank teachers, yet evaluators are expected to promote professional growth through nonjudgmental, formative supervisory processes. Teacher evaluation models are challenged to address the supervisory function to support educators’ professional growth and the accountability requirement to judge their competence for the purposes of continued employment decisions (Mead et al., 2012; Mette et al., 2017). Describing these contradictory goals, Danielson (2010) explained: “A system to ensure quality must be valid, reliable, and defensible—these are ‘hard-sounding’ qualities—whereas a system designed to promote professional learning is likely to be collegial and collaborative—these are much ‘softer-sounding’ qualities” (p. 37). For example, evaluators may be reluctant to take an authoritative approach and mandate teachers change their classroom practices, fearing a loss of political capital among teachers within their schools. As a result, leaders may inflate summative evaluation ratings and/or provide inaccurate positive feedback for the sake of maintaining peace and collegiality (Yariv, 2006). Some school leaders have evaluated teachers more leniently, fearing the risk of their own employment (Bridges & Groves, 1999) or declines in staff culture and relationships (Mitchell, 2011). Evaluators often softened the evaluation process with language to reduce defensiveness, but the reluctance of leaders to provide negative feedback has often perpetuated a status quo in which teachers are satisfied with marginal performance or may inaccurately believe their marginal performance is actually superior (Yariv, 2006).

**Professional development.** Improved teaching and learning practices require a clear process for improving teacher performance (Killion & Hirsh, 2011). Effective professional
development provides all teachers with time to collaborate, interact, and build trusting professional relationships with their colleagues (Copland & Knapp, 2006; Van Driel & Berry, 2012). In some states or districts, administrators may refer underperforming teachers to a peer assistance program, a teacher-led program to match experienced, competent educators with those needing assistance (Darling-Hammond, 2013).

Targeting professional development to the specific needs of teachers is crucial. Professional development may benefit from designs that differentiate the expectations based on experience, differentiating between the different performance expectations of novice and experienced teachers (Clayton, 2013). This alignment can provide a continuum of coordinated professional development extending throughout a teacher’s career (Darling-Hammond, 2014; Phillips, 2011). For example, alignment between new teacher induction programs and the district teacher evaluation process can help new teachers improve their practice at the crucial start of their career (Goldrick et al., 2013). Even with rigorous evaluation systems, financial and time resources must be devoted to building the capacity of teachers (Darling-Hammond, 2014).

**Terminating incompetent or unproductive personnel.** The summative evaluation rating is used for identifying teacher effectiveness differences and removing underperforming teachers who are less effective (Mead et al., 2012). Some researchers have expressed skepticism about the improvement of mediocre or ineffective teachers through staff development processes (Little & Miller, 2007). Administrators can work with teachers to implement improved instructional practices, but teachers may be slow or resistant to change (Murphy & Meyers, 2008). Therefore, improved hiring practices reduce the need remove ineffective and underperforming teachers later (Rose, English, & Finney, 2014). Hanushek (2009) argued for a one-time “deselection” to remove the lowest performing 10% of the teaching workforce. Rather
than taking time to improve teacher competence through professional development and training, the permanent removal of the lowest-performing staff may yield performance gains. According to Rivkin, Hanushek, and Kain (2005), “a one standard deviation increase in average teacher quality for a grade raises average student achievement in the grade by at least 0.11 standard deviations of the total test score distribution in mathematics and 0.095 standard deviations in reading” (p. 434).

**Unifying teachers and administrators in a common effort of educating students.** The final purpose of evaluations is to develop a shared understanding of effective teaching and learning practices. For leaders to supervise and evaluate teachers accurately, there must be a common understanding of teacher effectiveness (Danielson, 2010; Marzano et al., 2011). Having district-defined teaching standards creates a common, unified definition of quality teaching (Darling-Hammond, 2013; Marzano et al., 2011). Coherence is created when the evaluation process adheres to and reinforces these standards (Copland & Knapp, 2006). Standards may facilitate an “accountability relationship” (Holland, 2004, p. 7) between teachers and administrators, and this common language provides a framework to help evaluators hold professional conversations about student learning and classroom practices with teachers (Kraft & Gilmour, 2016). The growth of teaching models such as the Danielson (2013) teaching framework and Marzano’s teacher evaluation model (Marzano et al., 2014) aligned teacher evaluation systems to uniform standards, defining a common understanding of quality teaching and learning. In the next section, the research on defining teacher quality will be explored.

**Defining Teacher Quality**

Recent scholarship has focused on defining teacher quality through perceived classroom effectiveness. Darling-Hammond (2013) defined teaching quality as “strong instruction that
enables a wide range of students to learn” (p. 12) within the context and conditions where learning and teaching occur. Whitcomb and Rose (2008) defined an effective teacher as one whose teaching practices lead to improved student learning. Teaching is a complicated endeavor, and confounding variables limit the ability to prove causation between teaching inputs and student learning. Many complex factors influence student learning, and isolating single factors may result in misleading findings (Darling-Hammond, 2013). Some teaching inputs are difficult to measure, including a teacher’s effect on student social outcomes such as cooperation and work ethic (Goe et al., 2008). Research has attempted to identify variables to predict quality teachers, but children and schools are difficult to study—considering the challenges of researcher access and student mobility (Whitcomb & Rose, 2008).

The Bill and Melinda Gates Foundation (Gates Foundation) funded $696 million for teacher-related programs between 2008 and 2013, funding many teacher evaluation programs, including the Measures of Effective Teaching (MET) project. The foundation-supported research attempted to define and measure teaching quality through observations of over 3,000 teachers in an effort to capture the elements of teaching effectiveness (Phillips, 2011). This research determined the research-based teaching frameworks developed by Danielson (2013) and Marzano et al. (2014) provided a common terminology for describing quality teaching, and when combined with student feedback and student growth measures, trained evaluators could reliably measure teaching performance (Ruzek et al., 2011). Goe et al. (2008), writing for the National Comprehensive Center for Teaching Quality, presented this definition of effective teaching based on their analysis of empirical research on teaching:

1. Effective teachers have high expectations for all students and help students learn, as measured by value-added or other test-based growth measures, or by alternative measures.
2. Effective teachers contribute to positive academic, attitudinal, and social outcomes for students, such as regular attendance, self-efficacy, and cooperative behavior.

3. Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence.

4. Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness.

5. Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly for the success of students with special needs and those at high-risk for failure. (p. 8)

Research has challenged assumptions that a teacher’s age, certification, experience, and education level can predict teacher quality (D’Augustino & Powers, 2009). Although teachers’ unions have promoted experience as a measure of teaching quality, additional teaching experience beyond the early induction phase of a teacher’s career provides no additional benefit to improving student learning (Rivkin et al., 2005). One study of urban high schools estimated just 10% of the variation in teacher effects were attributed to teacher attributes such as pay, tenure, or experience (Aaronson et al., 2007). Jacob’s (2007) research in urban schools found few differences among the traditional credentials used to define quality teachers, finding “certified teachers are not consistently more effective than uncertified teachers, older teachers are not more effective than younger teachers, and teachers with advanced degrees are not more effective than those without such degrees” (p. 138). To increase professional accountability, NCLB included criteria to define a “highly qualified” teacher: the individual has a bachelor’s degree, holds a state teaching license, and has subject-matter expertise (NCLB, 2002). However, this reform was seen as largely bureaucratic and ineffective, failing to improve student learning (Mead et al., 2012). In the next section, research on the distribution of summative evaluation ratings will be explored.
Identification of Underperforming Teachers

Reports such as *The Widget Effect* (Weisberg et al., 2009) highlighted the disparity between student performance and summative evaluation ratings assigned to teachers in schools. Referred by some scholars as a “Lake Wobegon” effect, many evaluators have been criticized for rating most teachers as satisfactory or higher, inflating summative ratings as if all teachers were “above average” (Tucker, 1997; Zirkel, 2010). Tucker (1997) provided a detailed breakdown of teacher competency, explaining “the typical principal with a staff of 100 teachers identifies 1.53 incompetent tenured teachers per year and remediates 0.68 teacher, encourages 0.37 teacher to resign or retire, reassigns 0.29 teacher, and recommends dismissal for 0.10 teacher” (p. 1).

Research studies used a variety of methods, including survey research of principals, to determine an estimated number of underperforming teachers—those teachers identified for remediation or improvement. Bridges (1992) cited estimates that 5% of teachers were “incompetent” and harming student achievement. Lavely, Berger, and Follman (1992) conducted a meta-analysis of prior research studies to conclude 10% of teachers were “incompetent.” Tucker (1997) interviewed Virginia principals, concluding 5% of teachers in that state were underperforming but only .68% had undergone a remediation plan. Thompson (2006) surveyed principals in California, concluding 3.5% of California teachers were underperforming but only 0.70% were in a remediation plan.

In 2009, The New Teacher Project (TNTP) published a national report studying teacher evaluation in 12 low performing, urban school districts (Weisberg et al., 2009). In this national sample, approximately 99% of teachers were rated positively, using ratings such as good or great, or satisfactory or excellent. Officials in over half of these schools had not dismissed a tenured teacher during the 2-5 year survey period. This report was an influencing factor in the
creating of SB7, as three Illinois districts—Elgin, Rockford, and Chicago—were among the 12 districts studied. In these Illinois districts, 92.6% of teacher were rated “superior” or “excellent,” 7% were rated “satisfactory,” and only 0.4% were rated “unsatisfactory.” The report recommended adopting comprehensive evaluation systems, training evaluators, connecting evaluations to personnel functions, and creating dismissal policies (Weisberg et al., 2009).

Critics of the report, however, cited irregularities, noting one of the districts studied—Toledo, Ohio—had a robust program for eliminating underperforming teachers whose data were not included in the findings. The district had removed or remediated 12.9% of tenured teachers during 2008 and 8.5% of probationary teachers over a 5-year period (Thompson, 2010). In a follow-up study, Kraft and Gilmour (2017) analyzed statewide evaluation ratings data from 19 states, finding 2.7% of teachers received a rating of unsatisfactory performance. In follow-up interviews with principals from one district, the study found a discrepancy between the number to teachers perceived as underperforming (27.8%) and the number of teachers rated as underperforming (6.5%). Table 1 summarizes the actual ratings and estimated rates of teacher underperformance, as shared through these studies.

Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Teachers rated underperforming</th>
<th>Tenured teachers in remediation</th>
<th>Population surveyed</th>
<th>Estimates of underperforming teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois Report Card (ISBE, 2017d)</td>
<td>3.0%</td>
<td>n/a</td>
<td>Data from all Illinois public schools</td>
<td>n/a</td>
</tr>
<tr>
<td>Kraft and Gilmour (2017)</td>
<td>6.5%</td>
<td>n/a</td>
<td>Administrators surveyed in one large urban district</td>
<td>27.8%</td>
</tr>
<tr>
<td></td>
<td>2.7%</td>
<td>n/a</td>
<td>Data from 19 states</td>
<td>n/a</td>
</tr>
</tbody>
</table>

(continued)
Table 1 (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Teachers rated underperforming</th>
<th>Tenured teachers in remediation</th>
<th>Population surveyed</th>
<th>Estimates of underperforming teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weisberg et al. (2009)</td>
<td>0.4%</td>
<td>n/a</td>
<td>Administrators in Chicago, IL</td>
<td>7.5%</td>
</tr>
<tr>
<td></td>
<td>0.9%</td>
<td>n/a</td>
<td>Administrators in Rockford, IL</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>0.5%</td>
<td>n/a</td>
<td>Administrators in Elgin, IL</td>
<td>1.9%</td>
</tr>
<tr>
<td>Thompson (2006)</td>
<td>n/a</td>
<td>0.70%</td>
<td>California high school principals</td>
<td>3.5%</td>
</tr>
<tr>
<td>Tucker (1997)</td>
<td>1.1%-1.53%</td>
<td>0.68%</td>
<td>Virginia principals</td>
<td>5%</td>
</tr>
<tr>
<td>Bridges (1992)</td>
<td>n/a</td>
<td>n/a</td>
<td>Unknown</td>
<td>5%</td>
</tr>
<tr>
<td>Lavely et al. (1992)</td>
<td>n/a</td>
<td>n/a</td>
<td>Meta-analysis</td>
<td>10%</td>
</tr>
</tbody>
</table>

In November 2017, ISBE released the first statewide report of summative evaluation ratings—as required by SB7 legislation. Illinois public school districts reported the number of teachers earning an “excellent” or “proficient” rating during the 2016-2017 school year. This information was published on the Illinois Report Card website, allowing public access to this data for any public school district (ISBE, 2017d). Statewide, with 771 districts reporting, 97% of teachers were rated in the highest two categories. In the state’s largest school system, Chicago Public Schools, 89% of teachers were rated in the highest two categories. In analyzing these reports, a Chicago Tribune editorial praised the transparency but concluded, “Evaluations alone don’t drive superior performance. But honest appraisals give teachers feedback so they can improve. They may encourage the best teachers to stay and the worst to leave” (“What the report card says,” 2017, para. 11).

**Summative evaluation ratings distribution in Michigan.** Michigan presents an interesting case of evaluation policy reform. Michigan, a state with collective bargaining laws similar to Illinois, underwent similar reform efforts in 2009 and 2011, including changing to four-tiered summative evaluation rating system, with ratings of “highly effective,” “effective,”
“minimally effective,” and “ineffective” (Lenhoff, Pogodzinski, Mayrowetz, Superfine & Umpstead, 2017). Advocacy groups promoted the reforms as addressing teacher underperformance by identifying more underperforming teachers. Education Trust-Midwest pushed for broader feedback and a wider distribution in the evaluation scores, explaining, “When nearly all teachers are told they are doing well, expectations are lowered or remain ill-defined and teachers miss-out on the opportunities to help students learn” (Lenhoff, 2012, p. 3).

Data from two studies provided summative evaluation ratings from the first year of implementation. When Education Trust-Midwest studied 8,600 summative evaluation ratings from a sample of Michigan school districts during the 2011-2012 school year—the year after the new ratings system went into effect statewide—less than 1.2% of teachers were rated as minimally effective or not effective (Lenhoff, 2012). Data from the Michigan Department of Education supported this conclusion, finding less than 3% of teachers rated as minimally effective or not effective (Keesler & Howe, 2012). The Education Trust-Midwest report questioned these results as being too lenient and not representing the broad distribution for which they had hoped (Sawchuk, 2012). Yet, the Michigan Department of Education defended the results as “appropriate, particularly given that this is the first year of implementation” (Keesler & Howe, 2012, p. 9). These incongruent interpretations reflect the tenor of the teacher reform movement as internal stakeholders defend the status quo while external groups demand aggressive reforms with demonstrable results. Table 2 compares the distribution ratings from these two Michigan reports.
Table 2

Distribution of Ratings in Michigan after Year One of Four-Tier Evaluation (2011-2012)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly effective</td>
<td>23%</td>
<td>11.60%</td>
</tr>
<tr>
<td>Effective</td>
<td>75%</td>
<td>87.75%</td>
</tr>
<tr>
<td>Minimally effective</td>
<td>2%</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>Ineffective</td>
<td>&lt;1%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

In 2017, a Michigan study found 97% of Michigan teachers were rated in the two highest ratings categories. An analysis of statewide evaluation data from 2011-2012 and 2014-2015 found trends in the identification of underperforming teachers consistent with the earlier findings reported in *The Widget Effect*. The study concluded the stricter teacher evaluation policy reform in Michigan—a policy similar to SB7 and PERA in Illinois—failed to meet the policymakers’ objective to increase the identification of underperforming teachers (Lenhoff et al., 2017).

**Teacher dismissals in Illinois.** Illinois statutes have included a process for remediating and dismissing teachers underperforming since 1982. However, historical data are limited, as ISBE did not collect data on the number of teachers with underperforming ratings and/or remediation plans, nor data on teachers who were terminated or resigned. Therefore, Illinois researchers attempted to estimate the distribution of unsatisfactory ratings given to underperforming teachers. Scholars examined the number of teacher dismissals and/or remediations through analysis of dismissal hearings and survey research of school administrators. However, none of this data included teachers who resigned rather than face dismissal, those who agreed to financial settlements and confidentiality agreements in exchange for resigning, nor probationary teachers who were non-renewed (Henry, 2010; London, 1998; Seltzer, 1992; Thompson, 2010). Summarizing the difficulty of gathering data, London
concluded, “There is no way to determine the total number of tenured teachers were either dismissed or negotiated to leave the district” (p. 11).

Seltzer (1992) surveyed Illinois principals whose schools had one or more teachers under a remediation plan. First, data from ISBE indicated 92 districts had issued an unsatisfactory rating, but when contacted later, eight districts reported no one had received a remediation plan. Next, the study surveyed 61 principals in those identified districts. The principals reported 22 of their remediation plans resulted in teacher resignations while nine were dismissed.

Three studies examined teacher dismissal hearings in Illinois. Thurston (1990) reviewed 180 teacher dismissal hearings from 1976-1989. Over this 25-year period, an average of 12.3 teachers were dismissed annually through dismissal officer hearings. The review concluded Illinois lacked a common definition of incompetency—with each local board of education defining local standards of teacher underperformance. London (1998) studied 118 teacher dismissal hearing records from 1985-1994, finding an average of 11.8 teachers were dismissed annually through dismissal hearings. The study recommended ISBE should collect data on teacher dismissals. Henry (2010) examined teacher dismissal hearing records from 1990-2008. During that period, 219 tenured teachers were dismissed from 62 school districts after undergoing remediation, with 138 of those dismissed teachers coming from the Chicago Public Schools, averaging 11.5 teachers dismissed annually. Table 3 summarizes the Illinois teacher dismissal hearing data from these three studies.

Table 3

*Average Annual Teacher Dismissals from Dismissal Hearing Data in Illinois*

<table>
<thead>
<tr>
<th>Author</th>
<th>Time period</th>
<th>Years</th>
<th>Average annual dismissals</th>
</tr>
</thead>
</table>
Newspaper reporter Scott Reeder (2005a, 2005b)—a statewide reporter for the Small Newspaper Group, a publisher with local newspapers in Moline, Ottawa, and Kankakee, Illinois—mailed FOIA requests in 2005 to each of the 876 Illinois public school districts, requesting disclosure of information regarding remediation plans and teacher dismissals during a 10-year study period between 1995 and 2005. By using the FOIA procedure, Reeder achieved a 100% response rate, publishing his findings in the Illinois regional newspapers owned by Small Newspaper Group—including the Daily Journal (Kankakee, Illinois), The Times (Ottawa and Streator, Illinois), and Argus-Dispatch (Moline and Rock Island, Illinois). The study found few teachers beyond the probationary period are ever dismissed, consistent with findings that have been reported by Hanushek (2009). For both tenured and non-tenured teachers, the number of annual unsatisfactory ratings—as measured by the number of remediation plans—averaged one for every 930 teacher evaluations, with a statewide average of 51 unsatisfactory ratings annually. The study also found 83% of schools did not issue an unsatisfactory rating during the 10-year study period (Reeder, 2005a). In the next section, the history of evaluation policy and reforms in Illinois will be discussed.

**Illinois Evaluation Policy and Reforms**

Illinois has undergone five periods of teacher evaluation policy. In the first phase, prior to 1975, boards of education enjoyed broad freedoms in dismissing teachers, as state laws provided few teacher employment rights. The second period began in 1975 when Illinois passed Article 24-12, adding the hearing officer process to teacher dismissals when the dismissed teacher requested a hearing. In the third period, beginning in 1986, the Illinois School Reform Act 24A was passed. The law added additional processes for remediation and teacher dismissal (London, 1998). This law was passed partly in response to negative attention from *A Nation at Risk*, a
federal report that questioned student achievement in America’s schools. The law included a three-tiered evaluation rating system: “excellent,” “satisfactory,” and “unsatisfactory.” A tenured teacher who received an “unsatisfactory” rating would complete a 180-day remediation plan intended to help the teacher improve his or her performance (Seltzer, 1992). The fourth period of Illinois reform began in December 1997 when Illinois revised the school code, reducing the remediation period to a 90-day plan (Henry, 2010). In either case, a teacher could be dismissed only after they failed to meet the requirements of the remediation plan.

The fifth and current period of evaluation reform began in 2010 when, driven by the option to apply for RttT funding, the Illinois legislature passed the Performance Evaluation Reform Act (PERA). The law required all schools to include student growth in their summative teacher evaluations by 2015-2016 (ISBE, 2015). The law was passed during a veto session in January 2010 to meet the deadline for legislative changes required to qualify for a federal RttT grant. Under RttT, many decisions about implementation were left to state and local policymakers (Hill, Charalambous, & Kraft, 2012). Illinois applied for RttT funding, and by the conclusion of third funding round, 19 states were awarded RttT grants, including $43 million for Illinois’ third round award (Howell, 2015; Regenstein, 2011).

PERA reformed teacher evaluation in many significant ways. Prior to PERA, Illinois law called for three evaluation rating categories: “excellent,” “satisfactory,” and “unsatisfactory.” An “unsatisfactory” rating trigged a remediation plan. PERA created a new evaluation rating category—”needs improvement”—that requires teachers to complete a PDP, changing the previous evaluation ratings under 24A from three categories to four. In addition, PERA included new requirements to evaluate teachers using objective student growth data, with various timelines and requirements leading to full implementation in 2015-2016 (ISBE, 2015). Besides
to goal of increasing the likelihood the state would win RttT funding (Howell, 2015), advocates hoped the law would increase student achievement through the identification of underperforming teachers and their resulting improvement or dismissal.

In June 2011, the Illinois legislature followed-up with passage of a second reform package. Under SB7, the Illinois Superintendent of Schools may take action on removing teacher certificates for incompetence. Incompetence is defined as having two or more unsatisfactory evaluations within a 7-year timeframe. In determining an incompetence case, the State Superintendent must consider the remediation steps undertaken, the time between evaluations, and whether the evaluations occurred prior to PERA. A state certification board makes the final determination regarding revocation. The law also broadened the definition of who could evaluate teachers. Anyone who passes the required training could evaluate teachers starting on September 1, 2012. These qualified evaluators could include teacher-leaders, assistant principals, deans, department chairpersons, and others besides the principal. However, schools were warned against expanding the practice whereby a member of the teachers’ union would be evaluating a peer member of their union, complicating potential documentation of underperforming teachers in the future (ISBE, 2015). Table 4 summarizes the major components and changes found in PERA and SB7 legislation.

Table 4

Summary of Components of PERA and SB7 Legislation

<table>
<thead>
<tr>
<th>PERA (enacted January 2010)</th>
<th>SB7 (enacted June 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Evaluations using four rating categories</td>
<td>• Reduction in Force incorporates teacher evaluation into position lists to determine order of layoffs</td>
</tr>
<tr>
<td>• Non-tenured teachers evaluated annually, tenured teachers evaluated every two years</td>
<td>• Potential for tenure portability when teachers with switch employers (continued)</td>
</tr>
</tbody>
</table>

(continued)
Table 4 (continued)

<table>
<thead>
<tr>
<th>PERA (enacted January 2010)</th>
<th>SB7 (enacted June 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Annual evaluations for tenured teacher who received the needs improvement or unsatisfactory ratings</td>
<td>• Potential for school boards to grant tenure after three years for a teacher rated excellent in each of their first three years</td>
</tr>
<tr>
<td>• Professional Development Plan (PDP) triggered when teacher receives a needs improvement rating</td>
<td>• Streamlined teacher dismissals, including the new Optional Alternative Evaluation Dismissal (OAED) Process</td>
</tr>
<tr>
<td>• Dismissal when teachers received two unsatisfactory ratings within 36 months</td>
<td>• Required school board member training</td>
</tr>
<tr>
<td>• Mandatory training for evaluators</td>
<td>• Defines incompetence as two unsatisfactory evaluations within a seven-year period and permits the State Superintendent to revoke a teaching license for that reason</td>
</tr>
<tr>
<td>• Expands the ability to evaluate teachers beyond licensed administrators to teachers who complete training</td>
<td>• State Superintendent may act to require professional development for a teacher</td>
</tr>
<tr>
<td>• Student growth scores incorporated into evaluation ratings for both teachers and principals</td>
<td>• Survey of teaching and learning conditions must be conducted</td>
</tr>
<tr>
<td></td>
<td>• Statewide collection of teacher and principal evaluation ratings data</td>
</tr>
<tr>
<td></td>
<td>• Mandatory public disclosure of collective bargaining proposals prior to initiating a labor strike</td>
</tr>
</tbody>
</table>

*Note.* Adapted from Illinois Education Association [IEA] (2011) and ISBE (2015)

The lobbying for both PERA and SB7 included multiple stakeholders, including advocacy groups that hoped the final bills would result in more aggressive reforms (Regenstein, 2011). While advocacy groups played an influential role, some critics have questioned the influence of advocacy groups, especially regarding politically controversial subjects (Lubienski, Weitzel, & Lubienski, 2009; Scott et al., 2009). During lobbying efforts for PERA and SB7, advocacy groups such as Stand for Children and Advance Illinois infused campaign donations, lobbying efforts, and public media spending into public relations campaigns advocating for evaluation reforms. Numerous Chicago philanthropists, including current Illinois governor Bruce Rauner, real estate investor Sam Zell, businessperson Penny Pritzker, and other well-known business leaders backed these reforms with financial and political support (Long, 2011).
Some questioned whether the reforms were watered down and unlikely to produce the desired outcomes, citing the complex union entanglement and local decisions on how to implement many aspects of the reforms. For example, a district’s PERA student growth plan must be negotiated with the local teachers’ union, creating a potential obstacle for rigorous implementation. Education blogger Alexander Russo (2011) questioned the motives of PERA reform when he asked, “Is this model legislation developed through a model process, as is being claimed, or is it reform ‘lite,’ setting the bar extremely low and letting everyone declare victory without doing any of the really hard work?” (para. 4). Following the first year of reform implementation in their district, Chicago Public School teachers raised questions about the effectiveness of PERA and described their student growth performance tasks as easy to “game” (Jiang, Sporte, & Luppeco, 2015). Russo argued the reforms were compromises with teacher’s union who were faced with more unfavorable legislative proposals that could have become law. Karen Lewis (2011), president of the Chicago Teachers Union, issued a statement supporting the passage of SB7, explaining how the final bill avoided harsher anti-union reforms such as the repeal of collective bargaining rights and the right to strike. In the next section, the research on factors that influence evaluators when issuing summative ratings will be reviewed.

**Factors Influencing Identification of Underperforming Teachers**

Factors beyond teaching quality explain how evaluators are influenced in making decisions about teacher performance ratings. Kimball and Milanowski (2009) found ratings were influenced by three factors: skill, motivation, and context. Skill was described as the competence and training of the administrator, motivation—the desire and will to act, and context—the characteristics of the school environment in which the administrator was working. The local context includes procedural, personal, and logistical barriers influencing the decision making of
teacher evaluators. The next section will describe several of these factors and their influence on evaluators when determining summative ratings.

**Workload and time constraints.** Leaders’ workloads presented barriers to implementing evaluation policy reform. Evaluation reforms required leaders to make time for evaluations—classroom observations, documentation, and conferencing—in addition to meeting the other needs of students and staff (Donaldson & Papay, 2014; Drake et al., 2016). Leaders may lack the time in their workday to conduct effective evaluations given the myriad demands and responsibilities of their positions (Hallinger & Murphy, 2013; Kraft & Gilmour, 2017). This time pressure leads to increased work hours for leaders who are hindered by myriad interruptions and non-instructional responsibilities (Lavigne & Chamberlain, 2017). Leaders can reevaluate their management responsibilities and distribute non-instructional decision-making and managerial duties to others—redirecting their time toward activities to improve student achievement (Donaldson & Papay, 2012). Furthermore, when leaders need to address an underperforming teacher, the amount time needed for documentation, observations, and professional development may cause evaluators to issue a higher rating than was earned (Kraft & Gilmour, 2017). Leaders simply may “give-up” and approach evaluations with indifference (Fullan, 2014).

**Low confidence in effectiveness of remediation.** School leaders are skeptical or indifferent on the effectiveness of remediation efforts. Scholars also have questioned their ultimate potential to improve student learning, which is the primary goal of teacher evaluation (Murphy, 2013). When surveyed in 1992, Illinois principals who had written a remediation plan felt the process was effective (Seltzer, 1992), and among Wyoming principals, 61% found remediation plans to be effective (Range, Scherz, Holt, & Young, 2011). However, other
research has found principals have low confidence in remediation plans and prefer to bypass the formal evaluation process to handle improvement activities outside the evaluation process (Kraft & Gilmour, 2016). Teachers can meet the minimal requirements of the plan, but the plan may not result in sustained improvement (Holland, 2004). Principals may wait for the teacher to retire rather than engage a process whose outcome is uncertain or likely to foster change (Reuland, 2012). Principals may also use “harassing supervision practices” (Stoelinga, 2010, p. 58) to force a voluntary separation. One study from the Chicago Public Schools found examples of principals who used techniques such as reassigning teachers to undesirable grades or assigning undesirable classroom locations that, for example, required teachers to climb several flights of stairs each day (Stoelinga, 2010).

When remediation leads to short-term improvement, principals perceive these changes to be temporary (Reuland, 2012). Improving existing teachers takes time; meanwhile, students may suffer from poor instruction by an underperforming teacher (Hallinger & Murphy, 2013). One Illinois superintendent expressed his cautious reservations that remediation could lead an “unsatisfactory” teacher toward improvement, explaining, “I wouldn’t say these teachers became good teachers, but after they finished remediation, they weren’t doing any harm” (Reeder, 2005b, para. 26). Because of their low perceptions of the effectiveness of remediation, many principals have low motivation to pursue remediation of unsatisfactory teachers. Some feel they are unlikely to find a better teacher than the one who needs improvement (Kraft & Gilmour, 2017). Danielson (2010) admitted the limited potential of evaluation to improve student learning when she wrote, “principals need to devote time to the evaluation process—despite the fact that it often produces few benefits” (p. 39).
Challenge of legal issues. Myriad legal rules are daunting for administrators, including labor-management structures such as collective bargaining rights and union representation (Menuey, 2007; Reuland, 2012), tenure rules (Dandoy, 2012), laws and school code (Reuland, 2012), and the time requirement to document unsatisfactory performance (Dandoy, 2012; Menuey, 2007). These legal rights vary from state to state, and in some states, vary based on the collective bargaining agreements in effect within local districts. Labaree (2010) summarized this challenge:

The contract with the teacher union makes it extraordinarily difficult to fire a teacher after she has completed an initial probationary period of three years or so. American teachers enjoy a form of tenure guarantee that, though weaker than the tenure rights of American college professors, is much stronger than in nearly any other occupational setting. To fire a teacher after being tenured is so onerous in its requirements for documentation, due process, and battles with the union that most principals don’t even try. (p. 126)

Administrators cited the lack of funds and budgetary constraints to process and defend legal cases for dismissal. From 2001-2005, Illinois school districts engaged in dismissal litigation spent an average of $219,504.21 in legal costs for each tenured teacher dismissal proceeding (Reeder, 2006). In some states, dismissed teachers continue to earn their salary during the legal processes of hearings, litigation, and appeals (Bireda, 2010). In an era of shrinking funding for schools, administrators must weigh the financial costs against the other outcomes—including student achievement losses from an underperforming teacher (Rivkin et al., 2005).

The pressures of legal concerns influenced administrators when deciding whether to pursue remediation, especially the additional time needed to write and implement remediation plans (Range et al., 2011; Reuland, 2012). The burden of proof in dismissal cases is high, and lack of time to carry out these mandates was cited by 74% of California high school principals as a significant deterrent to dismissals (Thompson, 2006). Balancing an existing workload with
time needed to write the remediation plan and follow the necessary legal rules creates an enormous burden on the principal. One Illinois middle school principal described participation in the remediation process as a significant “investment” (Reuland, 2012, p. 164) of his time. Most schools simply lack the capacity and time needed to evaluate all teachers frequently (Chait, 2010). One New York attorney estimated principals would spend 10-15% of their work time managing a remediation, and summarized the frustration of administrators by asking,

what rational person would invest 15 percent of her time for two years just to get the teacher back in her building? It is taken as a given that when it comes to incompetent tenured teachers, the best you can do is to tell them to go to another school. (Levin, Mulhern, & Schunck, 2005, p. 18)

Reeder (2005a) concluded it was “almost impossible” to remove an underperforming teacher in Illinois. Reeder interviewed educational leadership professor Richard Manatt, who articulated his interpretation of Illinois’ evaluation system:

What we are looking at in Illinois is scandalous. Evaluations have turned into ritualistic endeavors–everybody receives a good evaluation. Teacher unions like it because it protects their members. Most administrators like it because they don’t like to rock the boat and that’s the sort of evaluation they get from their superintendents. (Reeder, 2005a, para. 37)

Before removing a low-performing staff member through the evaluation process, administrators might explore alternative dismissal options. For example, Whitaker (1999) recommended using other documentation strategies to build the legal case of insubordination prior to moving to remediation. As opposed to teaching performance that may be considered remediable and not subject to termination, insubordination is irremediable and subject to dismissal in most states. They may use “harassing supervision” (Stoelinga, 2010, pp. 57-58) tactics such as pressure, assignment changes, and voluntary separation agreements to remove an underperforming teacher. Novice principals are more likely to pursue the formal remediation process, while experienced principals are more likely to pursue other methods to achieve a
similar outcome (Reuland, 2012). While the barriers of remediation and dismissal in Illinois are significant, Henry (2010) stated, “(Illinois) school districts that are willing to accept the time and financial responsibility associated with dismissing a tenured teacher can effectively do so” (p. 125).

**Low confidence in correlations between evaluation ratings and student achievement.** Research has studied the linkage between student achievement outcomes and the evaluation rating of the teacher(s) and their students. For example, scholars raised numerous challenges to the validity and reliability of value-added measures to accurately account for student differences and questioned their role in evaluation reforms and teacher accountability (Amrein-Beardsley, 2008; ASA, 2014; Newtown, Darling-Hammond, Haertel, & Thomas, 2010). According to Popham (2007), “there is ample evidence that, instead of improving instructional quality, ill-conceived accountability programs can seriously diminish it” (p. 147). Additionally, even with rigorous evaluator training and specific evaluation criteria tools, evaluation ratings and student achievement may not correlate. Evaluators may have difficulty distinguishing practice from performance (Lenhoff et al., 2017).

**Politics and organizational culture.** The literature finds political barriers and school cultural factors have influenced administrators in their efforts to address teacher underperformance (Bridges & Groves, 1999; Reuland, 2012). The evaluation process sometimes creates a negative relationship between teachers and the administrators. The administrators may face negative perceptions from the staff and/or community as some persons may support and advocate for the underperforming teacher (Whitaker, 1999).

The value school districts place on teacher dismissal correlate with the likelihood a principal will move towards teacher dismissal, suggesting the importance of district culture to
address underperformance (Thompson, 2006). Among Illinois middle school principals, 20% reported being told or having perceived they could not place a teacher on a remediation plan at one time in their career (Reuland, 2012). District leadership must support administrators in their work to confront teacher dismissal and support the work of their evaluators during teacher evaluation conflicts (Bridges & Groves, 1999). As one study of the political climate in California schools concluded that

> few principals anywhere risk losing their jobs if they are less than forthright with a poor-performing teacher. However, if they evaluate a teacher negatively, the teacher may sow seeds of discontent and lower morale in the school. Thus, principals have incentives to bite the marshmallow, not the bullet; by doing their jobs properly, they heighten their vulnerability. (Bridges & Groves, 1999, p. 331)

**Lenient ratings to avoid staff conflicts.** Maintaining the dual roles of leader and evaluator, some evaluators may be motivated to rate leniently (Kraft & Gilmour, 2017). When they perceive the school culture could be disrupted, administrators may rate some teachers with higher ratings (Manatt & Daniels, 1990). They seek to maintain positive relationships with staff and feel lower evaluation ratings will affect those relationships, especially with veteran teachers (Kraft & Gilmour, 2017). Evaluators may also fear their own job security when conflicts with staff develop (Bridges & Groves, 1999). Lenient ratings may reduce or prevent conflict with teachers’ unions who might intervene on behalf of a teacher who is upset or disappointed by their evaluation rating (Mitchell, 2011).

Even teacher evaluations in high-stakes situations—such as merit pay—are subject to this leniency “because principals have to work with the teachers after their evaluation is complete, principals may still tend to inflate ratings . . . to maintain collegiality” (Kimball & Milanowski, 2009, p. 63). This sentiment may reflect the dichotomy of principals who are expected to motivate and lead their staff while also carrying out supervisory and evaluative duties (Holland,
Principals are expected to use the evaluation process to lead school improvement in their schools but may question the influence of their work on actually improving achievement (Mead et al., 2012; Murphy, 2013).

Supervision, the formative nature of the evaluation process, may contribute to leniency (Haefele, 1993). Rather than viewing evaluation as strictly summative, some evaluation systems are used as a professional development process (Marzano et al., 2011). The resulting written feedback given to teachers tends to be positive and descriptive—with minimal constructive criticism (Bernstein, 2004; Yariv, 2006). Although principals may perceive this process as helping the teacher and building positive relationships, some teachers perceive it as unhelpful:

At best, the evaluations I received were a mild form of feedback: I’m doing okay, I’ll have a job next year, I’ve met my professional development goals, and so forth. At worst, the current evaluation system was used as a tool for compliance and as a threat so that teachers did not ‘make waves’. As a union president, I have found that teacher evaluation is implemented inconsistently from school to school and even among the same administrators within the same school. (Bernstein, 2004, pp. 80-81)

Supportive and trusting relationships between leaders and teachers are crucial for productive school culture (Fullan, 2014), but leaders are challenged to maintain positive relationships with the same teachers who earned low performance ratings on evaluations (Mead et al., 2012).

**Personal discomfort with difficult staff conversations.** Developing a school culture that embraces challenging conversations about teaching performance in an important element of teacher improvement (Mead et al., 2012). However, leaders were uncomfortable confronting teachers about their low performance and avoided giving negative feedback (Kraft & Gilmour, 2017; Yariv, 2006). Evaluators may soften the language of feedback to reduce their discomfort (Danielson, 2010; Yariv, 2006). The cultural expectation of teachers’ autonomous independence in their classrooms may contribute to this reluctance, as “leaders avoid interfering with the work of teachers. They buy compliance with the currency of autonomy” (Murphy, 2013, para. 5).
Gaps in the Literature

Because the legislative changes are relatively recent, the literature is limited on the effect of PERA and SB7. It is too early to find peer-reviewed studies in educational journals. The literature regarding Illinois’ evaluations relies upon limited evidence from adjudicated remediation cases under the previous 24A laws that did not include the “needs improvement” category and the PDPs enacted under PERA (Henry, 2010).

The state of Illinois historically has not collected data on evaluation ratings. The research studies that attempted to measure the number of remediation plans or teacher dismissals were limited by incomplete data. Newspaper reporter Reeder (2005a, 2005b) generated the most thorough evaluation analysis available in Illinois, having collected data from 100% of Illinois schools from a 10-year timespan. Despite the national attention and professional awards his reporting earned, the work was not peer-reviewed and the results were published in the mass media instead of scholarly publications; however, his work has been cited in the literature (see Chait, 2010; Mitchell, 2011). The limited scholarly work on Illinois evaluations comes mostly from dissertation research (Henry, 2010; London, 1998).

Quantitative analysis of current evaluation data in Illinois will fill a significant research gap (Henry, 2010; London, 1998; Reeder, 2005a, 2005b; Seltzer, 1992; Thurston, 1990) and inform our understanding of the conditions under new policies, including the distribution of evaluation ratings. The SB7 law included provisions for collecting this data:

Section 24A-20(c) of the School Code requires that “districts . . . submit data and information to the State Board on teacher and principal performance evaluations and evaluation plans in accordance with procedures and requirements for submissions established by the State Board. Such data shall include, without limitation, (i) data on the performance rating given to all teachers in contractual continued service, (ii) data on district recommendations to renew or not renew teachers not in contractual continued service, and (iii) data on the performance rating given to all principals.” (ISBE, 2015, p. 32)
ISBE (2015) explained why this provision was included in the law:

Since SB7 required all teachers, including non-tenured teachers, to be rated in accordance with the four rating categories, ISBE will collect data on performance rating for all teachers. In the state’s application for State Fiscal Stabilization Funds (as part of the American Recovery and Reinvestment Act), the state had to commit to collecting data on the performance component, the student growth component, and the final summative performance evaluation rating.

It is nevertheless important to note that, pursuant to Section 24A-20(a)(1) of the School Code, the data collected may only be publicly reported in a manner whereby no teacher or administrator can be personally identified. (p. 32)

These data were supposed to be published by schools in their annual school report cards:

Section 10-17A of the School Code [105 ILCS 5/10-17A] requires ISBE to annually publish a report card. As part of that report card, ISBE is to publish the combined percentage of teachers rated as proficient or excellent in their most recent evaluation. State law prohibits the personal identification of any individual teacher. (ISBE, 2015, p. 32)

However, by the 2015-2016 school year, the only evaluation data collected was for the RttT school districts and school districts identified in lowest 20% of student performance (ISBE, 2016).

**Conceptual Framework**

Two conceptual frameworks informed and framed this study: education policy implementation theory (Honig, 2006) and micropolitics of personnel evaluation (Bridges & Groves, 1999). Changing policies require educators to study policy implementation “to uncover their various dimensions and how and why interactions among these dimensions shape implementation” (Honig, 2006, p. 14). These outcomes of policy implementation have been defined as

the product of interactions between polices, people, and places—the demands specific policies place on implementers; the participants in implementation and their starting beliefs, knowledge, and other orientations toward policy demands; and the places or contexts that help shape what people can and will do. (Honig, 2006, p. 2)
Therefore, research on policy implementation must extend beyond the structural policy itself and examine the localized conditions and processes in which the policy succeeded or failed. First, the policy itself may influence implementation (Malen, 2006). For example, Hill (2006) found the language of policy influenced implementation of mathematics reform, especially when terminology lacked precision. Actors interpreted different meanings of the policy’s language, resulting in implementation that was misaligned to the policy’s intent. Second, people who are involved with policy implementation have great influence on the policy’s success or failure (Malen, 2006). For example, Spillane (2008) found significant differences in policy implementation within and across school districts. In his study, school- and district-level leaders within the same district communicated different ideas about policy, resulting in mixed implementation results with varying degrees of fidelity and intensity observed at each school. Finally, the location and local context influences policy implementation (Malen, 2006). For example, Marsh and Crocker (1991) found local context influenced implementation of middle school reforms in eight California middle schools. Local conditions such as labor relations, leadership effectiveness, and collegial trust influenced the successes and challenges of implementation.

Policy implementation is influenced by political challenges that accompany change, and these challenges occur throughout implementation (Malen, 2006). Politics has been defined as “decisions related to the allocation of values for a given society or social organization; that is who gets what, when, and how” (Blase & Blase, 2002, p. 7). To examine the politics occurring in local contexts, micropolitics describes the public and private transactions that occur between actors who seek power and influence over decisions (Kirst & Wirt, 2009). Micropolitics appears often in the education literature to describe factors influencing decision making at the local level.
As democratic organizations, schools are localized venues for political debate and lobbying by actors who influence decision making and policy outcomes (Brosky, 2011; Malen, 2006).

Although researchers have focused on how state and local politics influence education policy outcomes, the study of micropolitics in education examines the “conditions, if any, various education policies get implemented and work” (Honig, 2006, p. 2) in the local schools. For example, Marsh (2012) examined the development of site-based, performance compensation policies in the New York City schools. Local school committees implement a system of incentive pay for educators and administrators based on student performance. The study found political actors—educators and administrators—used strategies to suppress conflict and hide or manipulate information. As a result, most schools implemented quality in pay rather than the differentiated pay intended by policymakers. Micropolitics research also examines how power is used to achieve goals in schools. Blase (1992) surveyed 836 educators who perceived their principals as open and effective. The study found these principals used a wide range of control and empowerment strategies including exchange, rewards, principal’s visibility in the school, and empowerment to influence teachers to meet the principal’s goals. Grissom, Kalogrides, and Loeb (2015) studied the micropolitics of assigning students to teachers and classrooms in one urban school district. The study found veteran educators used their influence and power to determine which students were assigned to their classrooms. Veteran educators received fewer disadvantaged students in their classrooms, resulting in inequalities when the neediest students were assigned to less experienced educators.

Political influence can impede the evaluation process (Blase & Blase, 2002; Bridges & Groves, 1999). To examine this influence, Bridges and Groves (1999) defined a micropolitical framework to describe the interactions of educational politics in teacher evaluation. First,
decisions interact with actors who possess different levels of access. These actors have interests
that evolve into goals and objectives. Next, conflicts may occur as actors use the power of their
influence. In addition, actors may form coalitions with other actors who share common interests.
Finally, actors use strategies to reach their goals, and outcomes are the decisions made.

Education evaluation involves decisions regarding the ground rules, the procedures, and the
evaluation itself, including the summative evaluation rating. Bridges and Groves noted:

> We argue that each type of personnel evaluation decision involves multiple actors with
> access to these decisions and particular interests and goals. When the interests of actors
> conflict, the opposing parties mobilize their power, form coalitions, and pursue strategies
to affect the outcome. The outcomes of the conflict reflect the interests of those actors
with the greatest power and most effective strategies. These outcomes may limit the
discretion that can be exercised regarding future decisions, or they may lead to
unanticipated negative consequences that trigger a new round of decisions. (p. 322)

Recent policy trends—such as Illinois’ evaluation reforms—have changed the balance of power
in evaluation relationships, becoming more evenly distributed between the teacher and evaluator
roles. In examining this trend, Bridges and Groves (1999) posed five questions for investigating
these changes.

1. How have the ground rules, procedures, and performance evaluations changed over time?
2. Whose interests have predominated or been frustrated by these changes?
3. How have the sources and levels of power, coalitions, and strategies of various actors
   stimulated these changes?
4. To what extent have the excesses and abuses of dominant actors stimulated these
   changes?
5. What political forces, if any, eventually lead to ground rules, procedures, and
   performance evaluations that more adequately balance the interests of prime beneficiaries
   (i.e., parents and students) in quality and the interests of employees in job security and
   fair treatment? (Source: Bridges & Groves, 1999, p. 336)

The interaction of education policy implementation theory and micropolitics of personnel
evaluation framed this study. This interaction is embedded in Kirst and Wirt’s (2009) definition
of micropolitics: the “formal and informal transactions that shape policy implementation” (p. 203). It is important to understand local policy implementation through micropolitics, and researchers have used micropolitics as a lens to describe the complexities of policy implementation (Marsh, 2012). As just one dimension of policy implementation, politics can influence the success or failure of policy reforms (Kirst & Wirt, 2009). Malen (2006) asserted that politics is always present in policy implementation, proposing a framework that examines “reciprocal relationships” between policy implementation and politics, with the challenge for researchers “to unpack how politics creates policy, and to uncover how policy creates politics” (p. 85). An example, Malen described how policy actors may pushback during implementation to alter the intended outcome of the policy. Marshall and Scribner (1991) described the policymaking process as a central concept of micropolitics, focusing on “the processes of arriving at decisions, policies, regulations, and decrees emanating from power relationships and conflict situations” (p. 349). Micropolitics is important in policy implementation research because conflicts can reveal the values of leaders and stakeholders in the organization, not just “managerial conflicts to overcome” (Flessa, 2009, p. 346). Schools are often considered to be positive (or neutral) institutions in which the presence of conflict is not normal; addressing the micropolitics of schools may expose difficult social justice or equity issues that were hidden previously but need to be illuminated (Datnow, 2001; Flessa, 2009). Therefore, in order to understand the implementation of teacher evaluation reform in this study, a micropolitics lens was used to illuminate the decisions, actors, and outcomes of policy implementation.

**Conclusion**

Policy reform advocates promoted teacher evaluation reform in Illinois as a mechanism to improve student learning through the identification and improvement of underperforming
teachers (Regenstein, 2011). However, teachers and leaders hold low confidence in the ability of evaluation reforms generally to influence changes in teaching and learning practices (Donaldson, 2012). As one scholar bluntly explained, “The reality is that there is very little evidence to suggest that it [teacher evaluation] has done anything to improve schooling” (Murphy, 2013, para. 3). The literature suggests that few underperforming teachers are identified through the formal evaluation process (Bridges, 1992).

The implementation of education policy is affected by local context, including the complex interactions of policies, people, and places (Malen, 2006). Micropolitics of personnel evaluation provide one lens for examining the interactions of local actors who may influence evaluators when identifying underperforming teachers (Bridges & Groves, 1999). This chapter presented an overview of leadership for learning, teacher evaluation, and the policies and practices of teacher evaluation in Illinois. In addition, the conceptual frameworks of education policy implementation (Honig, 2006) and micropolitics of personnel evaluation (Bridges & Groves, 1999) were described as frameworks to guide the study.
Chapter 3

Methodology

The ability of policymakers to measure the outcomes of policy reforms is vital to the development of future policies. Because policies are implemented at the local level, research that includes local data and local policy actors can illuminate policy implementation and micropolitics. This chapter describes the research questions for the study in addition to the research design, population, development and validation of the instruments, data collection, ethical considerations and validity, and the data analysis procedures.

Research Questions

The following research questions guided this study:

**Research question 1.** To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools?

**Research question 2.** How have micropolitical factors influenced principals in the identification of underperforming educators in Illinois since the implementation of teacher evaluation reforms?

Research Design

This study utilized a two-phase, explanatory sequential mixed methods design to answer two research questions. A visual model of the study design is included in Appendix B. The literature uses many terms that are synonymous with mixed methods—including mixed methodology, synthesis, and multimethod—to describe a research design that incorporates the use of both quantitative and qualitative data (Creswell & Plano Clark, 2011). Mixed methods is common in policy research as an approach to examine a phenomenon while also measuring the
outcomes and implementation of policy (Burch & Heinrich, 2016). For this study, the mixed methods design was chosen for the purpose of complementarity—using multiple methods to gain a deeper understanding of different facets of a single phenomenon (Greene, 2007).

In the quantitative phase, data were collected to determine whether the policy reform caused a significant difference in the identification of underperforming teachers. In the qualitative phase, data were collected to explain whether micropolitics influenced the implementation of the policy reform. The explanatory sequential variant was chosen so participants in the qualitative phase with knowledge of the mechanisms behind the quantitative data could explain and illuminate the quantitative data (Cameron, 2009). This design also integrated the quantitative and qualitative results into a cohesive analysis to strengthen and justify the overall conclusions of the study (Ivankova et al., 2006).

Mixed method studies reflect worldviews unique to each phase of the study. Although the literature includes debate on the nature and role of worldviews (Greene, 2007), four commonly discussed worldviews are postpositivist, constructivist, participatory, and pragmatist. The quantitative study reflects a postpositivist worldview, focusing on a narrowly defined, empirical data set regarding teacher evaluation and the identification of underperforming teachers (Teddle & Tashakkori, 2009). The qualitative portion reflected a constructivist worldview. The analysis sought to understand the data on teacher evaluation by gathering credible descriptions from multiple evaluators. The epistemology of the constructivist worldview co-constructs the reality with participants (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2009). These postpositivist and constructivist worldviews supported the explanatory sequential mixed methods design for this study. The quantitative data provided empirical evidence of teacher evaluation
outcomes while the qualitative data illuminated the influence of micropolitics on the implementation of teacher evaluation reform.

The unit of analysis in this study was principals who evaluate teachers. The study examined the micropolitical factors that influence their decisions in teacher evaluation.

**Population, Site Selection, and Participants**

The site for both phases was the state of Illinois. This state was chosen as an interesting case of teacher evaluation policy implementation (Regenstein, 2011). In addition, prior Illinois studies on teacher dismissal conducted by Henry (2010), Reeder (2005a, 2005b), and Weisberg et al. (2009) were available for historical comparison.

**Quantitative phase population.** In the quantitative phase, the population ($N = 859$) included all public elementary, high school, and unit districts in Illinois identified from the Directory of Educational Entities provided to the researcher in September 2017 (ISBE, 2017a). These entities are required to evaluate teachers according to the regulations for all public schools in Illinois. Upon request, ISBE provided the email addresses for all district superintendents and principals. The accuracy of these addresses is dependent upon the timely uploading of updated administrator contact information by the school district. Because the administration of 859 electronic questionnaires was feasible, the entire population of public school superintendents identified for this study was contacted by email to complete the questionnaire. Superintendents were chosen because, as leaders of their school districts, they have access to the data and administrative support staff needed to gather the requested information. The study did not require superintendents to work in the district during the years studied, and respondents could choose to provide partial data.
Ideally, all 859 Illinois superintendents would have replied to the questionnaire request. However, the questionnaire required data collection and time resources from the district to complete it, which may have discouraged superintendents from participating. Overall, 183 superintendents initiated the online questionnaire while 91 superintendents completed it. Six superintendents declined consent. Upon visual analysis, two questionnaires included data that did not conform to expected norms and were excluded from the study. Therefore, the final sample population for the quantitative phase was 89 school district superintendents responding, representing 10.36% of the 859 districts statewide.

**Qualitative phase population.** To define the population for the qualitative interview phase, a four-step analysis was used (Robinson, 2013). First, criteria were established to identify participants whose experiences are most likely to provide them the knowledge to illuminate the research questions (Palinkas et al., 2013). The population consisted of currently practicing principals in Illinois public schools who worked in roles as evaluators of educators for at least 7 years. Because the study sought participants whose experiences could illuminate the research questions, potential participants must have identified one or more teachers for a remediation plan, a PDP, or dismissal during their Illinois public school careers. The size of the qualitative sample for interviewing was established at 20 participants.

Third, a sampling strategy for the quantitative phase was selected. Purposeful criterion-iSampling was used to identify participants who met the predetermined criteria for inclusion in the study (Palinkas et al., 2013). Expert judgment was used to identify the participants to be invited for interviews, seeking respondents would provide the most insight (Robinson, 2013). In addition, a diverse pool of participants was identified to increase the generalizability of the study sample, representing diverse demographics, including gender, race/ethnicity, geographic regions
of the state, and district demographics (Teddlie & Tashakkori, 2009). Finally, the methods for recruiting and identifying participants were established. First, during the qualitative phase, superintendents offered suggestions for school leaders who would meet the criteria for the qualitative phase and whose experiences may provide rich data for the study. In addition, all principals in Illinois public schools were invited by e-mail to review the criteria and volunteer for consideration as a participant (Appendix F).

Telephone screening interviews were conducted with 28 volunteers who responded via email. This screening interview determined if volunteers met the criteria for inclusion and which cases had the greatest potential for rich and thick description relevant to the qualitative research question (Appendix G). From these screeners, 21 participants were interviewed by telephone or in person between December 6, 2017 and January 16, 2018. During one interview, it was learned one participant did not meet the criteria for the study, and this interview was excluded from the study and replaced with another participant. Therefore, the final number of interviewees was 20 principals. One participant served a dual role as a principal/superintendent and was included in the study. Pseudonyms and school district names were chosen by the participants or assigned by the researcher.

Each interview participant’s gender, race/ethnicity, years of evaluating educators in Illinois public schools, and their current roles were obtained during screening interviews (Table 5). Information on district size-type and region was obtained from ISBE (2017b; 2017c). Participants’ evaluation experience in Illinois public schools ranged from 7 years to 18 years (M = 12.15 years). Fourteen participants were from large school districts, four from medium-sized districts, and two from small districts. Eight participants were from elementary districts,
one from a high school district, and 11 from unit districts. Participants included 10 males and 10 females, representing equal gender representation.

Participants represented districts throughout the state of Illinois. The six districts of the ISBE Statewide System of Support were used to classify the geographic region of Illinois represented by the participants (Figure 1). Eleven participants (55%) represented Area 1 (Northeast), including Cook County and Chicago Public Schools, four participants (20%) represented Area 2 (West Central), no participants represented Area 3 (West Central), three participants (15%) represented Area 4 (East Central), one participant (5%) represented Area 5 (Southwest), and one participant (5%) represented Area 6 (Southeast) with one participant (5%). (Table 5). Although respondents were more heavily represented in Area 1, this region is the most populous in Illinois and has the most school districts.

![Figure 1. ISBE Statewide Systems of Support (https://www.isbe.net/Pages/Statewide-System-of-Support-Fiscal-Agents.aspx)](https://www.isbe.net/Pages/Statewide-System-of-Support-Fiscal-Agents.aspx)
Table 5

Participants in Qualitative Phase

<table>
<thead>
<tr>
<th>Participant</th>
<th>School district</th>
<th>Gender</th>
<th>Race/ethnicity</th>
<th>Years as evaluator in Illinois</th>
<th>District size-</th>
<th>SSOS area/region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aidan O’Brien</td>
<td>Irving</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>11</td>
<td>Large-elementary</td>
<td>1</td>
</tr>
<tr>
<td>Amanda Ashbee</td>
<td>Sunfields</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>9</td>
<td>Medium-unit</td>
<td>2</td>
</tr>
<tr>
<td>Ann Keaton</td>
<td>Napa River</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>17</td>
<td>Large-unit</td>
<td>5</td>
</tr>
<tr>
<td>Charlie Ramirez</td>
<td>Churchill</td>
<td>Male</td>
<td>Hispanic/Latino White</td>
<td>10</td>
<td>Large-elementary</td>
<td>1</td>
</tr>
<tr>
<td>Corey Hammer</td>
<td>Ferndale</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>14</td>
<td>Large-unit</td>
<td>4</td>
</tr>
<tr>
<td>Diana Legend</td>
<td>Rosedale</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>17</td>
<td>Large-unit</td>
<td>4</td>
</tr>
<tr>
<td>Dustin Ross</td>
<td>St. Paul</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>8</td>
<td>Large-unit</td>
<td>2</td>
</tr>
<tr>
<td>Emma Moore</td>
<td>Gregson</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>7</td>
<td>Large-elementary</td>
<td>1</td>
</tr>
<tr>
<td>Eric Graves</td>
<td>Founders</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>9</td>
<td>Medium-unit</td>
<td>2</td>
</tr>
<tr>
<td>Frank Steele</td>
<td>Bayview</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>17</td>
<td>Large-unit</td>
<td>1</td>
</tr>
<tr>
<td>Glen Tucker</td>
<td>Crater Bay</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>17</td>
<td>Large-unit</td>
<td>1</td>
</tr>
<tr>
<td>Harold Cooper</td>
<td>Kappel</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>9</td>
<td>Large-high school</td>
<td>1</td>
</tr>
<tr>
<td>Jeffrey Shannon</td>
<td>Highlane</td>
<td>Male</td>
<td>Not Hispanic/Not Latino Asian</td>
<td>18</td>
<td>Large-unit</td>
<td>4</td>
</tr>
<tr>
<td>Jordyn Harris</td>
<td>Hillman</td>
<td>Female</td>
<td>Not Hispanic/Not Latino Black or African-American</td>
<td>10</td>
<td>Medium-elementary</td>
<td>1</td>
</tr>
<tr>
<td>Melody Roberts</td>
<td>Pittman</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>10</td>
<td>Large-unit</td>
<td>1</td>
</tr>
<tr>
<td>Michelle Hale</td>
<td>Austin</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>13</td>
<td>Large-unit</td>
<td>1</td>
</tr>
<tr>
<td>Molly Martin</td>
<td>Logan</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>8</td>
<td>Small-elementary</td>
<td>6</td>
</tr>
<tr>
<td>Rebecca Silver</td>
<td>Jasper</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>10</td>
<td>Large-unit</td>
<td>2</td>
</tr>
<tr>
<td>Stephen Pierce</td>
<td>Maple Ridge</td>
<td>Male</td>
<td>Not Hispanic/Not Latino White</td>
<td>14</td>
<td>Medium-elementary</td>
<td>1</td>
</tr>
<tr>
<td>Vickie Wells</td>
<td>Abbidale</td>
<td>Female</td>
<td>Not Hispanic/Not Latino White</td>
<td>15</td>
<td>Large-unit</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Pseudonyms used. Some data from ISBE (2017b; 2017c).
The diversity of interview participants closely paralleled the overall diversity of all principals in Illinois, as reported by White and Brown (2010). Table 6 lists race and ethnicity data for participants in comparison to all principals in Illinois.

Table 6

*Participant Race and Ethnicity*

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Number of participants</th>
<th>Percentage of participants</th>
<th>Statewide in 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>1</td>
<td>5%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Not Hispanic/Not Latino</td>
<td>19</td>
<td>95%</td>
<td>94.1%</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0</td>
<td>0%</td>
<td>&lt; 0%</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Black or African-American</td>
<td>2</td>
<td>10%</td>
<td>14.8%</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>0</td>
<td>0%</td>
<td>n/a</td>
</tr>
<tr>
<td>White</td>
<td>16</td>
<td>80%</td>
<td>80.1%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>1</td>
<td>5%</td>
<td>3.9%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>10</td>
<td>50%</td>
<td>52%</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>50%</td>
<td>48%</td>
</tr>
</tbody>
</table>

*Note.* Data from 2008 study by White and Brown (2010).

**Development and Validation of the Questionnaire**

Prior to data collection in the quantitative phase, a pilot study with three districts was conducted to test the reliability and validity of the survey instrument (Appendix E). In addition, a practicing Illinois school law attorney was consulted about the questionnaire items and process for data collection.

**Data Collection**

The data for this study were collected in two phases as summarized in the data collection matrix in Appendix A. In phase one, data were obtained from two sources. First, in September 2017, a Freedom of Information Act (FOIA) request was emailed to the Illinois State Board of Education. FOIA requests are often used in mixed methods research to identify the topics for further study in the second stage (Savage & Hyde, 2014; Walby & Larsen, 2012). Electronic
records were requested to obtain data on public school districts and the contact names and email addresses of district superintendents and school principals (Appendix D). Second, during October 2017, requests (Appendix C) were emailed to the superintendents of 859 Illinois public school districts. These notices informed and alerted participants that an online questionnaire would be emailed the following week. The email included an attachment with a copy of the questionnaire, allowing the superintendent or other staff member to begin gathering the data prior to receiving the questionnaire. Data included the number of teachers placed on remediation plans, teachers assigned to PDPs, and teacher dismissals for the 11 school years between 2006-2007 and 2016-2017. In addition, survey participants were asked for the school year when the district initiated PDPs triggered by a “needs improvement” rating required by Section 24A-5(h) of the School Code following the passage of PERA. One week after the worksheet was emailed, a link to the online questionnaire was sent by electronic mail. Survey Monkey was used to communicate with participants and collect data for the online questionnaire. One week later, a follow-up email was sent requesting superintendents who had not completed the questionnaire to do so. Data from responding districts were supplemented with data from ISBE (2017b). By collecting data back to the 2006-2007 school year, this study helps to fill a gap in the literature on Illinois teacher dismissals since 2008 (Henry, 2010) and extend the data collected by Reeder (2005a, 2005b). Additionally, the study was strengthened by analyzing data before and after the implementation of policy reforms.

Information from the quantitative phase was explored further in the qualitative phase. During December 2017 and January 2018, in-depth semi-structured interviews were conducted with 20 participants to explore the micropolitical factors that influenced principals when evaluating teachers (Appendix K). Semi-structured interviews provide in-depth examination of
participants as they share their individual experiences and stories in their own words (Merriam & Tisdell, 2016; Seidman, 2013). The interview protocol (Appendix I) was designed to address the questions for further study of micropolitics of personnel proposed by the framework of Bridges and Groves (1999). The questions were open-ended, allowing the participants to reflect on and reconstruct their experiences rather than simply recall memories (Seidman, 2013). The interview lengths ranged from 18 minutes to 70 minutes ($M = 36$ minutes). Two in-person interviews were conducted, and 18 were held via telephone. The interview protocol contained a list of common questions and potential follow-up questions, allowing me to respond and interact with the participants’ responses (Merriam & Tisdell, 2016).

Each interview was recorded digitally to preserve the original spoken words of the participant. The recordings were transcribed by a professional transcription service to capture the accurate verbatim text of the participants. The audio recordings were deleted once the transcripts were verified. A copy of the themes and findings was provided to each participant for member checking, allowing participants to verify the inferences and interpretations (Seidman, 2013). Follow-up questions were emailed to all 20 participants to illuminate and explain the quantitative results. Responses were received from 12 participants and reported in the integrated findings section.

**Ethical Considerations and Validity**

The study was approved by the Institutional Review Board of the University of Illinois (Appendix J). Informed consent was obtained from all participants prior to interviews, as confidentiality is desirable when the topics are controversial or when identifying information potentially could harm the participants (Creswell, 2014; Yin, 2014). Approved consent forms clearly stated the rights of participants as human subjects in research, including the right to
discontinue their participation at any time (Appendix H). All case information was converted to pseudonyms and other fictitious identifiers will be included to protect the identity of the participants (Yin, 2014). Without these protections, gaining access to interview participants would have been difficult (Flessa, 2009). All individually identified information was kept confidential.

Although critics of mixed methods approaches cite inherent difficulties in maintaining a rigorous design that uses both quantitative and qualitative methodologies (Bogdan & Biklen, 2007), these challenges are mitigated by collecting and analyzing all data using rigorous procedures to ensure validity and reliability (Ivankova, 2014). Therefore, this study applied several strategies to ensure the validity and reliability of the findings. First, in mixed methods studies, analyzing data from multiple sources is one mechanism to confirm the accuracy of the data and findings (Creswell, 2014). In this study, data on underperforming teachers was collected separately in the quantitative phase, then further explored by the use of qualitative interviewing with evaluators separately in phase two. Validity is strengthened when data are collected independently of one another and when the phenomenon studied—in this instance, teacher evaluation—is consistent in both phases (Greene, 2007). Second, the coded results and emergent interview themes were shared with the participants to confirm if their viewpoints were accurately represented. Third, a peer debriefer reviewed the coding and interview analysis to confirm their accuracy and increase the validity of the findings. Finally, rich narration provided detailed descriptions to increase the validity of the findings (Lincoln & Guba, 1985).

Data Analysis

In the quantitative phase, the data were analyzed using quantitative data analysis procedures as described by Creswell and Plano Clark (2011). First, the data were prepared for
analysis using software—Microsoft Excel 2013 for the Windows computer platform. This preparation included aligning data from the questionnaires with data collected on the district from ISBE, coding data with numeric values, checking the data set for entry errors, calculating any new variables for data to be combined for analysis, and maintaining a codebook within the software to list and describe the variables. Second, I visually reviewed the data, prepared a descriptive analysis, and inspected the data for any interesting trends or distributions. Third, the data were analyzed to illuminate interesting findings and patterns. To begin this analysis, descriptive tables and graphs helped organize the data to show the frequency counts of educators who (a) started a PDP, (b) started a remediation plan, and (c) were dismissed from their positions. Aggregate data from 2006-2011 and 2011-2017—representing the years before and after the required implementation of PDPs—were analyzed to determine whether the identification of underperforming teachers was significantly different following the implementation of reforms. Because the total number of teachers represented in the population changed in each year of this analysis, ratios were calculated to show the number of teachers represented by each PDP, remediation, or dismissal. Fifth, the results were interpreted to illuminate the research questions. These findings were compared against other findings in the literature regarding the identification of underperforming teachers in Illinois (e.g., Henry, 2010; ISBE, 2017d; Reeder, 2005a, 2005b; Weisberg et al., 2009). Finally, the findings were checked for validity through consultation with the peer debriefer.

In the qualitative phase, the data were analyzed using the qualitative analysis procedures described by Creswell and Plano Clark (2011). First, interviews were transcribed by a professional transcription service. The transcriptions were uploaded and cataloged for analysis using computer software—HyperRESEARCH version 3.7.3 for the Windows computer platform.
Second, I read through all transcripts to form a general impression of the data. Within the software, a codebook maintained a database of codes that evolved during data analysis. Third, the data were analyzed using a categorical approach to data analysis, with the narrative text reorganized and categorized into codes for easier comparisons (Teddle & Tashakkori, 2009). These codes guided me to identify the “salient themes, recurring ideas or language, and patterns of belief that link people and settings together” (Marshall & Rossman, 2011, p. 214). Fourth, I explained the results through discussions of the themes in the next chapter with a table to summarize the results. The oral speech recorded during the interviews was edited for written presentation to preserve the dignity and confidentiality of the participants and to avoid any sensitive or vulnerable issues (Seidman, 2013). Fifth, the results were interpreted to illuminate the research questions. This interpretation was guided by my personal perspective and the conceptual frameworks of education policy implementation theory (Honig, 2006) and micropolitics of personnel evaluation (Bridges & Groves, 1999) presented in Chapter 2. The interpretation identified the major lessons from the narrative. Finally, the data and results were checked for validity through member checking. Participants were provided copies of the descriptions of the themes and results and given the opportunity to provide comments on their accuracy (Creswell, 2014).

The results of the quantitative and qualitative phases were integrated during the interpretation phase of the study to discuss the results of the entire study (Creswell & Plano Clark, 2011). As a sequential explanatory design, this section explained how the qualitative data explained the quantitative data, as well as how the overall research question is addressed (Ivankova et al., 2006). The study used a systematic process to ensure the validity of the meta-inferences (Ivankova, 2014). First, participants for the qualitative phase were selected to ensure
they met the identified criteria. Second, the phase two follow-up questions were informed by the results from phase one. Finally, I observed the interaction between the qualitative and quantitative phases. Data from follow-up questions to illuminate the quantitative findings were analyzed and reported as integrated findings. I monitored data throughout the study, formulated tentative themes, and continually revisited themes throughout data collection and analysis. The participants were invited to review the overall research findings at the conclusion of the study.

Conclusion

This chapter described the methodology for a mixed methods study to examine the influence of micropolitics on the implementation of teacher evaluation reforms in Illinois. In the quantitative phase, Illinois public school superintendents were surveyed to collect data on the number of educators identified as underperforming. In the qualitative phase, 20 principals were interviewed to determine the role of micropolitics on the implementation of evaluation policy reform and to expand upon the quantitative findings. In the next chapter, the results of each phase will be discussed separately, followed by a discussion of the integrated results.
Chapter 4

Results

This mixed methods study examined the phenomenon of teacher evaluation, focusing on how micropolitics have influenced the implementation of teacher evaluation reforms in Illinois. The following research questions guided this study:

1. To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools?

2. How have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms?

First, this chapter provides a description of the quantitative data collected for this study and the findings of the relevant descriptive statistics analysis to address the first research question. Second, the results of the qualitative phase will be described to address the second research question. Finally, the results of the qualitative phase will be integrated to more fully explain the data collected in the quantitative phase.

**Research Question 1: To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools?**

Data regarding PDPs, remediation plans, and educator dismissals were collected from 89 respondents, which represented 10.36% of Illinois’ 859 public school districts. In this section, the results of this questionnaire are described with narrative description and visual charts and graphs.

**Respondent demographics.** Data on student enrollment and educator counts were obtained from ISBE (2017b). For the purposes of this study, “educator” includes all certificated employees in non-administrative appointments (e.g., teachers, counselors, librarians, social workers, psychologists, deans, instructional coaches, etc.) who are evaluated under the Illinois
state teacher evaluation requirements. Based on enrollments during the 2016-2017 school year, respondent school districts comprised 5.82% of total student enrollment in Illinois \((N = 118,137)\). Based on 2016-2017 educator counts (ISBE, 2017b), respondent school districts employed 6.37% of educators in Illinois \((N = 8,113)\). Thus, although respondents represented 10.36% of districts statewide, respondents represented a smaller percentage of statewide student enrollment and educator count. Two factors explain this difference. First, the three largest school districts in Illinois did not complete the questionnaire. Second, the state of Illinois contains many small, rural districts with small enrollments and small staff sizes (Table 7).

Table 7

Respondents for Quantitative Phase

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Respondent districts</th>
<th>All districts statewide</th>
<th>Percentage of respondents among all districts statewide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Districts</td>
<td>89</td>
<td>859</td>
<td>10.36%</td>
</tr>
<tr>
<td>Student enrollment</td>
<td>118,137</td>
<td>2,000,730</td>
<td>5.90%</td>
</tr>
<tr>
<td>Educator count</td>
<td>8,113</td>
<td>127,310</td>
<td>6.37%</td>
</tr>
</tbody>
</table>


Table 8 compares student enrollment of respondent districts with overall Illinois public school district enrollments. Respondents included a higher proportion of districts with smaller enrollments than the statewide averages. In addition, the student enrollment range was much narrower than district enrollments across the state, attributed to the absence of the three largest school districts in the respondent group.
Table 8

*Enrollment Comparisons Between Respondents’ Districts and All Illinois Public School Districts*

<table>
<thead>
<tr>
<th>Districts</th>
<th>Range (number of students)</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents</td>
<td>56-12,246</td>
<td>1,326.54</td>
<td>462</td>
</tr>
<tr>
<td>Statewide Total</td>
<td>40-340,668</td>
<td>23,374.69</td>
<td>727</td>
</tr>
</tbody>
</table>


When comparing the enrollments of responding districts to overall public school district enrollments in Illinois, district type and size were considered. First, Illinois school districts are organized by three grade-level configurations: unit (K-12), elementary (K-8), or high school (9-12). Respondents included 44 elementary (49.44%), seven high school (7.87%), and 38 unit districts (42.70%). The distribution of responding districts by school district type generally was similar to the statewide distribution (Table 9). Next, the enrollment size of respondent districts by organization type was compared to overall Illinois district enrollments by type. ISBE classifies district sizes based on fall enrollment counts taken on September 30 of each school year. The 25% of districts with lowest enrollments are classified as “small,” the 25% with highest enrollments are “large,” and the remaining districts are “medium” (ISBE, 2017b). Among respondents, 26.97% represented small districts ($N = 24$), 58.43% represented medium districts ($N = 52$), and 14.61% represented large districts ($N = 13$). Compared to the statewide data, the sample population had fewer large districts and more medium-enrollment districts (Table 9).
Table 9

*Frequency of School District Types and District Size in Quantitative Phase*

<table>
<thead>
<tr>
<th>District type</th>
<th>Respondents</th>
<th>Respondent district distribution (%)</th>
<th>Statewide district distribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary (K-8)</td>
<td>44</td>
<td>49.44%</td>
<td>43.02%</td>
</tr>
<tr>
<td>High School (9-12)</td>
<td>7</td>
<td>7.87%</td>
<td>10.12%</td>
</tr>
<tr>
<td>Unit (K-12)</td>
<td>38</td>
<td>42.70%</td>
<td>44.53%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>District size</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>24</td>
<td>26.97%</td>
<td>25%</td>
</tr>
<tr>
<td>Medium</td>
<td>52</td>
<td>58.43%</td>
<td>50%</td>
</tr>
<tr>
<td>Large</td>
<td>13</td>
<td>14.61%</td>
<td>25%</td>
</tr>
</tbody>
</table>

Note: Using 2016-2017 data (ISBE 2017a, 2017b)

To examine the geographic distribution of the respondents throughout Illinois, the six ISBE Statewide System of Support areas were used (Figure 1). Area 1 (Northeast) included 28 respondents (31.46% of respondent group); Area 2 (Northwest), 14 (15.73%); Area 3 (West Central), 12 (13.48%); Area 4 (East Central), 14 (15.73%), Area 5 (Southwest), (6.74%); and Area 6 (Southeast), 15 (16.85%). Respondents represented all of the SSOS education areas/regions in Illinois. Although respondents were more heavily represented in Area 1, this region is the most populous in Illinois and has the most school districts.

**Professional Development Plans: Staggered implementation.** The 2010-2011 school year represented the first full school year of the PERA law requiring districts to implement a Professional Development Plan (PDP) in their evaluation plan for educators. The PDP is developed in collaboration with the board of education and the collective bargaining agent representing educators. Districts could begin implementing these plans following the passage in PERA in January 2010 (ISBE, 2015). Respondents were asked to indicate which school year
they implemented the PDP requirement of PERA. The rollout of PDP implementation occurred over several years in school districts (Table 10). Only 3.37% of districts implemented in the first year \((N = 3)\) and an additional 5.62% implemented in the second year \((N = 5)\). By the 2016-2017 school year, 84.27% of districts had implemented the PDP requirement of PERA \((N = 75)\). However, 13 respondent districts indicated they have not implemented the required PDPs into their evaluations, representing 14.60% of respondent districts—even though PDPs are state-mandated for all Illinois public school districts.

Table 10

<table>
<thead>
<tr>
<th>Year Implemented</th>
<th>Respondents</th>
<th>% of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-2011</td>
<td>3</td>
<td>3.37%</td>
</tr>
<tr>
<td>2011-2012</td>
<td>5</td>
<td>5.62%</td>
</tr>
<tr>
<td>2012-2013</td>
<td>14</td>
<td>15.73%</td>
</tr>
<tr>
<td>2013-2014</td>
<td>15</td>
<td>16.85%</td>
</tr>
<tr>
<td>2014-2015</td>
<td>16</td>
<td>17.98%</td>
</tr>
<tr>
<td>2015-2016</td>
<td>13</td>
<td>14.61%</td>
</tr>
<tr>
<td>2016-2017</td>
<td>9</td>
<td>10.11%</td>
</tr>
<tr>
<td>Not implemented</td>
<td>13</td>
<td>14.61%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1.12%</td>
</tr>
</tbody>
</table>

Respondents provided data on the frequency of PDPs created with tenured educators; these tenured faculty members would have been rated as “needs improvement” in their summative evaluation ratings. During the first full year after PERA passage (2010-11), respondent districts implemented 6 PDPs with tenured educators; in 2016-2017, 24 PDPs were developed. The number of PDPs ranged from two (during 2009-2010) to 26 (during 2014-2015).
Respondents were not required to provide data for every year but were encouraged to provide whatever data they had available. Therefore, calculations were made to determine percentages and ratios based on the number of educators represented by respondents who reported data in each year. The number of educators reported to ISBE (2017b) was used. By calculating ratios in the sample population, estimates of frequency in the Illinois public school educator population can be calculated. These percentages, ratios, and statewide estimations are provided below (Table 11).

Table 11

| Tenured Staff Assigned a Professional Development Plan in Respondent Districts |
|-----------------|-----------------|-----------------|----------------------|-----------------------|
| School year     | Number of PDPs in respondent districts | Number of educators in respondent districts | Percentage of educators in PDP in respondent districts | Ratio of PDP:Educators in respondent districts | Estimated frequency of PDPs in all district statewide |
| 2009-2010       | 2                | 6,786            | 0.023%              | 1:3,393               | 39                  |
| 2010-2011       | 6                | 7,250            | 0.083%              | 1:1,208               | 106                 |
| 2011-2012       | 14               | 7,282            | 0.192%              | 1:520                 | 246                 |
| 2012-2013       | 13               | 7,834            | 0.166%              | 1:602                 | 216                 |
| 2013-2014       | 19               | 7,719            | 0.246%              | 1:406                 | 320                 |
| 2014-2015       | 26               | 7,973            | 0.326%              | 1:306                 | 423                 |
| 2015-2016       | 25               | 8,115            | 0.308%              | 1:324                 | 392                 |
| 2016-2017       | 24               | 8,113            | 0.296%              | 1:338                 | 377                 |

*Note.* Data on number of educators from ISBE (2017b).

When these data are charted in a line graph, omitting the 2009-2010 school year when the reform was passed mid-year, a declining trend line in the ratio of PDPs to Illinois public school educators is shown (Figure 2). However, districts implemented PDPs in a staggered fashion since passage of the legislation, gradually increasing with each subsequent year; thus, fewer districts implemented plans in the early years following PERA. By 2016-2017, when 75 (84.27%) of
respondent districts had implemented the PDPs, one PDP was implemented for every 338 educators working in the respondent school districts. As a result, a lower frequency of PDPs would be expected early in this reform because fewer districts had implemented their plans.

![Graph showing educators per PDP](image)

**Figure 2.** Ratio of educators to each PDP started.

Caution must be taken when interpreting these results, because these data may be skewed due to the absence of the three largest school districts in Illinois. During the 2016-2017 school year, for example, Chicago Public Schools (CPS) officials reported a higher frequency of educators identified with underperforming ratings than the overall statewide data showed (“What the report card says,” 2017). If the CPS schools experienced a higher frequency of identified underperformance in years prior to the 2016-2017, the data would not reflect educators statewide, but would be limited to downstate schools excluding the three largest districts.

**Remediation plans started.** Respondents were asked to report the number of educators starting a remediation plan during each school year between 2006-2007 and 2016-2017. Unlike
PDPs, remediation plans have been a required component in Illinois legislation since 1986 (London, 1998). Therefore, there were no concerns about staggered implementation of remediation plans. Respondents were not required to provide data for every year and were encouraged to provide whatever data they had available—including partial data. Therefore, calculations were made to determine percentages and ratios based on the number of educators represented by the respondents who reported data in each year. If a respondent did not report data, their district’s educator counts were excluded from calculations. The number of full-time educators reported annually to ISBE was used in these counts (2017b). By calculating ratios for respondents, estimates of frequency in all Illinois public school districts could be calculated. These percentages, ratios, and statewide estimations are provided (Table 12).

Table 12

<table>
<thead>
<tr>
<th>School year</th>
<th>Number of remediation plans implemented</th>
<th>Number of educators represented</th>
<th>Educators starting a remediation (%)</th>
<th>Ratio of remediation: educators</th>
<th>Estimated statewide frequency of remediation plans implemented based on ratio of remediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td>2</td>
<td>6,201</td>
<td>0.032%</td>
<td>1:3,101</td>
<td>41</td>
</tr>
<tr>
<td>2007-2008</td>
<td>1</td>
<td>6,361</td>
<td>0.016%</td>
<td>1:6,361</td>
<td>21</td>
</tr>
<tr>
<td>2008-2009</td>
<td>2</td>
<td>6,438</td>
<td>0.031%</td>
<td>1:3,219</td>
<td>41</td>
</tr>
<tr>
<td>2009-2010</td>
<td>2</td>
<td>6,786</td>
<td>0.029%</td>
<td>1:3,393</td>
<td>39</td>
</tr>
<tr>
<td>2010-2011</td>
<td>7</td>
<td>7,250</td>
<td>0.097%</td>
<td>1:1,036</td>
<td>124</td>
</tr>
<tr>
<td>2011-2012</td>
<td>7</td>
<td>7,282</td>
<td>0.096%</td>
<td>1:1,040</td>
<td>123</td>
</tr>
<tr>
<td>2012-2013</td>
<td>12</td>
<td>7,834</td>
<td>0.153%</td>
<td>1:653</td>
<td>199</td>
</tr>
<tr>
<td>2013-2014</td>
<td>2</td>
<td>7,719</td>
<td>0.026%</td>
<td>1:3,856</td>
<td>34</td>
</tr>
<tr>
<td>2014-2015</td>
<td>4</td>
<td>7,973</td>
<td>0.050%</td>
<td>1:1,993</td>
<td>65</td>
</tr>
<tr>
<td>2015-2016</td>
<td>8</td>
<td>8,115</td>
<td>0.099%</td>
<td>1:1,014</td>
<td>125</td>
</tr>
<tr>
<td>2016-2017</td>
<td>7</td>
<td>8,113</td>
<td>0.086%</td>
<td>1:1,159</td>
<td>110</td>
</tr>
</tbody>
</table>

Note. Data on number of educators from ISBE (2017b).
The aggregate number of remediation plans started by respondent districts ranged from one (during 2007-2008) to 12 (during 2012-2013). During the 11-year study period, a mean of 4.91 tenured educators were placed on a remediation plan annually. If this annual rate was applied statewide, the estimated statewide mean would be 83.79 tenured educators across all Illinois public school districts. However, examining the data since the passage of PERA, between 2009-2010 and 2016-2017, a mean of 6.25 tenured educators were placed on remediation plans annually. When data between 2009-2010 and 2016-2017 are charted in a line graph, the data show a declining trend line in the ratio of educators to remediation plans, indicating an increase in the number of tenured educators placed on remediation plans (Figure 4).

Figure 3. Ratio of educators for each remediation plan.
Because a remediation plan is triggered by an unsatisfactory summative evaluation rating, these results can be compared to evaluation data. Examining evaluation data from 1995-2005, Reeder (2005b) estimated 51 unsatisfactory ratings were issued statewide annually with a ratio of one educator rated unsatisfactory for every 930 evaluations of tenured and untenured educators. However, these data were reported prior to the PERA reforms that expanded the required evaluation categories to four, adding the needs improvement rating (ISBE, 2015). Between 2006-2007 and 2016-2017, 62 (69.66%) respondent school districts did not issue a remediation plan, while 27 (30.33%) districts had initiated one or more remediation plans. In comparison, Reeder (2005b) found 83% of districts did not issue a remediation plan from 1995-2005.

**All educator dismissals (tenured and non-tenured educators).** Respondents were asked to provide the number of educator dismissals (both tenured and non-tenured) for the school years between 2006-2007 and 2016-2017. Respondents were not required to provide dismissal data for every year and were encouraged to provide whatever data they had available—including partial data. Calculations were made to determine percentages and ratios based on the number of educators represented by respondents who reported data in each year. The number of educators reported to ISBE (2017b) was used. By calculating ratios in the respondent group, estimates of frequency in the population could be estimated. These percentages, ratios, and statewide estimations for dismissals of educators (tenured and non-tenured) are provided (Table 13).
### Table 13

**Educator Dismissals in Respondent Districts (Tenured and Non-tenured)**

<table>
<thead>
<tr>
<th>School year</th>
<th>Number of educator dismissals</th>
<th>Number of educators represented</th>
<th>Percentage of educators</th>
<th>Ratio of dismissals:educators</th>
<th>Estimated statewide frequency of educator dismissals based on ratio of dismissals</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td>4</td>
<td>6,201</td>
<td>0.065%</td>
<td>1:1,550</td>
<td>82</td>
</tr>
<tr>
<td>2007-2008</td>
<td>13</td>
<td>6,361</td>
<td>0.204%</td>
<td>1:489</td>
<td>269</td>
</tr>
<tr>
<td>2008-2009</td>
<td>19</td>
<td>6,438</td>
<td>0.295%</td>
<td>1:339</td>
<td>393</td>
</tr>
<tr>
<td>2009-2010</td>
<td>16</td>
<td>6,786</td>
<td>0.236%</td>
<td>1:424</td>
<td>312</td>
</tr>
<tr>
<td>2010-2011</td>
<td>18</td>
<td>7,250</td>
<td>0.248%</td>
<td>1:403</td>
<td>318</td>
</tr>
<tr>
<td>2011-2012</td>
<td>16</td>
<td>7,282</td>
<td>0.220%</td>
<td>1:455</td>
<td>281</td>
</tr>
<tr>
<td>2012-2013</td>
<td>19</td>
<td>7,834</td>
<td>0.243%</td>
<td>1:412</td>
<td>315</td>
</tr>
<tr>
<td>2013-2014</td>
<td>14</td>
<td>7,719</td>
<td>0.181%</td>
<td>1:551</td>
<td>236</td>
</tr>
<tr>
<td>2014-2015</td>
<td>23</td>
<td>7,973</td>
<td>0.288%</td>
<td>1:347</td>
<td>374</td>
</tr>
<tr>
<td>2015-2016</td>
<td>34</td>
<td>8,115</td>
<td>0.419%</td>
<td>1:239</td>
<td>533</td>
</tr>
<tr>
<td>2016-2017</td>
<td>35</td>
<td>8,113</td>
<td>0.431%</td>
<td>1:232</td>
<td>549</td>
</tr>
</tbody>
</table>

*Note. Data on number of educators from ISBE (2017b).*

When the data on tenured and non-tenured dismissals are charted in a line graph, the data show a slight declining trend line in the ratio of dismissals to educators in Illinois. The decline indicates a slight increase in the number of tenured educators dismissed during the 11-year study period (Figure 4). However, examining the trend since the implementation of PERA and SB7 reforms, the ratio decreased from 1:424 in 2009-2010 to 1:232 in 2016-2017. The small number of dismissals reported in 2006-2007 may be an outlier that skews the trend. When the linear trend line was calculated using just the years since PERA was passed in January 2010, the linear trend showed a decrease in the number of educators dismissed since PERA was enacted (Figure 5). This decline in educator dismissals possibly may be explained by the increase of PDPs as an additional lever for educator improvement during the same period.
Figure 4. Ratio of educators per dismissal (2006-2007 through 2016-2017).

Figure 5. Ratio of educators per dismissal since PERA (2009-2010 through 2016-2017).

Statewide evaluation data from 2016-2017. In November 2017, ISBE released statewide educator evaluation data from the 2016-2017 school year. This annual collection from
districts and the public release of evaluation data were mandated under SB7 (ISBE, 2015). For each school and district, the online *Illinois Report Card* (www.illinoisreportcard.com) listed the percentage of educators receiving an excellent or satisfactory rating on their evaluations during 2016-2017. Statewide, 97% of educators were rated as “excellent” or “satisfactory” during 2016-2017 school year (ISBE, 2017d).

Evaluation data reported by ISBE from respondent school districts in the quantitative phase was compared to the statewide data for all districts. Among respondent districts, ISBE reported 2016-2017 evaluation data for 78 respondent districts while 11 respondent districts did not have data available from ISBE. After removing the non-reporting districts, calculations were made to compare the respondent districts with the statewide data. In 2016-2017, 98.75% of educators \(N = 7,740\) in respondent districts were rated as “excellent” or “satisfactory,” with 1.25% of educators \(N = 98\) in respondent districts rated as “needs improvement” or “unsatisfactory.” Across all Illinois public school districts statewide, however, 3% were rated as “needs improvement” or “unsatisfactory” in 2016-2017. Data from Chicago Public Schools—the state’s largest school district—were included for comparison (Table 14).

**Table 14**

*ISBE Statewide Evaluation Data for 2016-2017*

<table>
<thead>
<tr>
<th>Rating</th>
<th>Number of educators</th>
<th>Reported percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educators rated as excellent or satisfactory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondent districts</td>
<td>7,740</td>
<td>98.75%</td>
</tr>
<tr>
<td>All Illinois public school districts</td>
<td>123,491</td>
<td>97%</td>
</tr>
<tr>
<td>Chicago Public Schools</td>
<td>17,383</td>
<td>89%</td>
</tr>
<tr>
<td>Educators rated as needs improvement or unsatisfactory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondent districts</td>
<td>98</td>
<td>1.25%</td>
</tr>
<tr>
<td>All Illinois public school districts</td>
<td>3,819</td>
<td>3%</td>
</tr>
<tr>
<td>Chicago Public Schools</td>
<td>2,149</td>
<td>11%</td>
</tr>
</tbody>
</table>

*Note.* ISBE (2017d).
Research Question 2: How have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms?

The qualitative phase examined the second research question, which involved the role of micropolitical factors in evaluators’ identification of underperforming teachers. Semi-structured interviews with 20 experienced Illinois public school evaluators were conducted. In the next section, the results of the interviews are presented to support four identified themes from the framework of Bridges and Groves (1999): (a) decisions about procedures of the evaluation plan; (b) decisions by evaluators; (c) actors and their access, interests, and power; and (d) outcomes of policy reforms. Each theme has subthemes described with narrative from the participant interviews (Table 15).

Table 15

Themes/Subthemes of Qualitative Phase

<table>
<thead>
<tr>
<th>Theme/Subtheme</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decisions about procedures of the evaluation plan</td>
<td></td>
</tr>
<tr>
<td>Joint committee collaboration: Local policy implementation</td>
<td>19 participants’ districts’ joint committees created the procedures; 1 was created by the administration.</td>
</tr>
<tr>
<td>Adoption of the Danielson teaching framework</td>
<td>19 participants’ districts adopted the Danielson teaching framework.</td>
</tr>
<tr>
<td>Selection of Type III assessments</td>
<td>4 participants described problems with Type III assessments.</td>
</tr>
<tr>
<td>Assessment quality and approval</td>
<td>9 participants described student growth assessments that required the minimum amount of growth or effort.</td>
</tr>
<tr>
<td>Procedures for student growth scores and the summative rating</td>
<td>15 participants described student growth procedures leading to inflated summative ratings.</td>
</tr>
<tr>
<td>Procedures influencing evaluators’ workload and focus</td>
<td>11 participants discussed procedures that increased evaluators’ workload.</td>
</tr>
</tbody>
</table>

(continued)
Table 15 (continued)

<table>
<thead>
<tr>
<th>Theme/Subtheme</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decisions by evaluators</strong></td>
<td></td>
</tr>
<tr>
<td>Interrater reliability</td>
<td>4 participants described concerns with interrater reliability in their districts.</td>
</tr>
<tr>
<td>Deferral of low ratings due to retirement/seniority</td>
<td>5 participants deferred low ratings due to the educators’ proximity to retirement or seniority.</td>
</tr>
<tr>
<td>Unwilling to issue a low rating</td>
<td>9 participants were unable or unwilling to issue a low rating to an underperforming educator.</td>
</tr>
<tr>
<td>Conversations about teaching and learning</td>
<td>9 participants described improved conversations about teaching and learning following the reform.</td>
</tr>
<tr>
<td>Reduction in Force (RIF)</td>
<td>5 participants utilized the RIF lever following the reform.</td>
</tr>
<tr>
<td>Intermediate lever of Professional Development Plans (PDP)</td>
<td>11 participants implemented a PDP following the reform.</td>
</tr>
<tr>
<td>Remediation plans</td>
<td>12 participants implemented remediation plans following the reform, with 3 participants experiencing remediation plans that resulted in teacher improvement.</td>
</tr>
<tr>
<td><strong>Actors and their access, interests, and power</strong></td>
<td></td>
</tr>
<tr>
<td>Union influence</td>
<td>11 participants described neutral to positive union influence when identifying underperformance.</td>
</tr>
<tr>
<td>School board influence</td>
<td>20 participants were not hindered by the school board in identifying underperformance.</td>
</tr>
<tr>
<td>Superintendent and central office influence</td>
<td>19 participants were not hindered by the superintendent or central office in identifying underperformance.</td>
</tr>
<tr>
<td>Parent and student influence</td>
<td>20 participants were not hindered by parents or students in identifying underperformance.</td>
</tr>
<tr>
<td>Educator influence and building climate</td>
<td>6 participants discussed how improvement plans ultimately improved building climate.</td>
</tr>
<tr>
<td><strong>Outcomes of policy reforms</strong></td>
<td></td>
</tr>
<tr>
<td>Value of improvement plans</td>
<td>20 participants answered affirmatively when asked if they would implement improvement levers again.</td>
</tr>
<tr>
<td>Mixed perceptions of policy reforms</td>
<td>20 participants felt educators had been treated fairly under the new evaluation plans, but perceptions of the policy’s outcomes were mixed.</td>
</tr>
</tbody>
</table>
Decisions about procedures of the evaluation plan. Decisions about procedures of teacher evaluation are the local “manner in which evaluations should be conducted” (Bridges & Groves, 1999, p. 322). The review of the literature described the regulations for teacher evaluation processes in Illinois (ISBE, 2015). In this section, decisions about the procedures of policy reform are explored. The subthemes included the work of the joint committee, adoption of the Danielson teaching framework, selection of Type III assessments, assessment quality and approval, procedures for student growth scores and the summative ratings, and procedures influencing evaluator workload and focus.

Joint committee collaboration: Local policy implementation. The rules of teacher evaluation reform delegated many decisions for local implementation to joint committees in each district. The joint committees, comprised of educators and administrators, were tasked with developing the procedures for student growth and teacher evaluation in their districts. Nineteen participants’ districts had active joint committees that created the procedures while one joint committee delegated the responsibility to the administration to create the procedures. Participants shared varied experiences as joint committees developed the procedures—ranging from collaboration to defensiveness.

Vickie described how her joint committee worked collaboratively to develop the student growth measures: “It was very collaborative. We had building representatives and different grade levels—and being a K-12 district, you’ve got everything from kindergarten, self-contained elementary, special education . . . all the way up to departmentalized in the high school.” Aidan concurred, describing his district’s positive process:

I’ve been really pleased with the conversations we’ve had. I don’t feel like its administrators want X and teachers want Y and we are at a gridlock or we’ve got to negotiate. I think it’s pretty open conversation.
There’s five administrators, five staff members across the district that are on it. I’m actually leading, I’m the chair of that committee right now. We have really good conversations. Throughout the year, we meet three times. . . . I don’t feel like anything has been politicized. It’s nice when one of the teacher members has this perspective on student growth or this perspective on the number of observations and another one has a very different perspective.

However, other participants described a less collaborative process where educators were entrenched to protect their interests or indifferent to the results. Amanda described the joint committee as a “defensive process” for educators. She found educators became entrenched protecting their interests against low evaluation ratings and potential job loss:

A very defensive process where the teachers are like, “We’re going to protect our own against this new system and these evaluators, to make sure that no one is going to just be out to get people to reduce expenditures in the district. To reduce either money that they’re spending on teachers’ salaries,” etc. That’s another way that the system just became a little more complicated.

Ann concurred, stating educators on the joint committee made the process “more complicated” to protect themselves from the possibility of receiving low ratings. Ann explained the educators and the teachers’ union wanted to protect their members from repercussions:

I think their concern was that their administrators could decide they just don’t like somebody and make it very personal and find a way to get rid of that teacher through this process. They made it a little difficult for me because I really couldn’t disagree with them. I guess from that standpoint, there was a little bit of, I don’t want to say contention, but kind of. We had lots and lots and lots of meetings and conversations before we came to an agreement. Then the document has been revised one time since then.

In contrast, Molly dealt with indifference from educators regarding the joint committee process. Educators were comfortable allowing the administration to create the plan—absent the involvement and input of the joint committee. Molly explained,

The problem is they trusted me so much to be fair. You know what I mean, because they knew me from previously. Even my parents in the community, they know me. They know I am going to be fair. I am not going to play one side or the other. So they trusted me to the point they were going to let me do it all, which really wasn’t fair, and that’s what I said. I said, “Number one, it’s not fair to me. And number two, if there is
something on there you don’t like . . . don’t even come back and tell me you don’t like it!”

*An evaluation framework clarifies expectations of effective teaching.* Among participants, 19 districts adopted the *Danielson Framework for Teaching* (Danielson, 2013) for their evaluation framework. Nine participants felt the district’s decision to utilize a framework, with its clearly articulated descriptions of teaching performance, made identifying underperforming educators easier. Evaluators easily could justify low ratings, supported by the language of the framework components.

Michelle felt using a teaching framework was a significant improvement over a checklist previously used in the Austin district, because the framework’s focus on teaching and learning made evaluations more objective. Jordyn discussed the depth of teaching behaviors described by the framework:

I think it’s easier, just because, again, the tool itself that most districts have to use, like Danielson, allows you to look at many aspects of an effective teacher in a detailed way. If you’re looking at planning, all these pieces that go into planning that, perhaps, you didn’t consider, before you maybe just say, “Okay, does the teacher even turn in plans?” and more of the compliance kinds of things. Some things of substance, but I think just being able to use a rubric to give you more of an in-depth look at what effective teacher looks like has been helpful.

In addition, Frank explained how the framework helped communication through specific criteria and expectations:

I think it provides some definite observable actions and the framework for people to be able to have conversations to improve, but also provides some specific behaviors and observations and data that people can look and see if things aren’t happening. I think being able to clearly articulate some of those expectations, and then reflect on what is happening in the classroom and beyond is very helpful to be able to communicate with staff and help them continue to grow.

Next, participants discussed that some educators need more training with the Danielson framework to understand their evaluation ratings and how teaching behaviors are measured.
Harold felt the framework made identifying underperforming educators easier, but described steps needed to improve educators’ understanding of the framework language:

I think it makes it easier to identify. I think what is required, though, is some more understanding, more calibration around that. Really unpacking what some of those words mean, some of those descriptors. What if schools and districts take the time to really unpack those rubrics with teachers around Danielson? It makes it a lot easier. I don’t think we’ve put enough time into helping teachers understand the rubrics.

Charlie agreed that educators lacked understanding of the evaluation framework language:

For the general population of teachers, they don’t understand it well enough. They think they do, but they really don’t. And they can’t get to that level, so they are not understanding that it’s a pretty high demand in order to get that excellent. They still want to get that distinguished or excellent rating, even though they’re not performing at that level on a regular basis, so they’re not quite understanding the rubric.

*Use of Type III student growth assessments.* Joint committees determined what assessment types were acceptable for use in their districts. Four participants cited the decisions to use teacher-created Type III assessments as easy for teachers to demonstrate student growth.

First, the pre- and post-test model inherently creates growth. Second, most Type III tests were single-response question formats, rather than open-ended, constructive response assessments.

Charlie explained:

It creates a system where if teachers are creating their own stuff—let me give you this test in French, let’s see what you know, which is going to be zero to almost nothing, and now I’m going to teach you French. Okay, now take the same test. Ah, amazing, you grew!

Later, Charlie expanded on his experiences with student growth assessments:

The student growth is not really giving us much of any value. We use teacher-created assessments, so they’re all pretty much Type III. Some are considered Type II because they’re (given) here and at our sister school or department-wide or grade level wide. But, in general, they’re created by teachers and they’re graded by teachers and then they’re scored by teachers and reported by teachers. And you (the evaluator) basically say “thank you!” for that information, and pretty much everybody’s getting either “meets” or “exceeds” in that component (student growth).
In Jeffrey’s school, the union favored using teacher-created and teacher-graded Type III assessments. He shared, “I don’t necessarily think that creates the most rigorous of assessments.”

Harold described how evaluation procedures in Kappel school district gave educators “total autonomy” over the student growth process:

In our district—and even in my previous district—we decided to go to the two Type III assessments for student growth. The teachers pretty much have total autonomy around what those measures are going to be and the feedback. Their partners who are looking at the data tend to be other teachers. That’s how it has been set up . . . where administration has been pulled away from looking at that data. Teachers are looking at it, they’re looking at the student growth, and they’re making the determination with very little administrative input into how well they’re doing or how well their students are growing. They’re giving themselves higher scores because they have complete autonomy about what it is that they want to measure in terms of their growth.

The administration has been pulled away from being able to say, “Hey, these are the measures we want to take a look at, or want to use a Type 1 or Type 2 measure.” They’re basically sandbagging it, to be honest with you.

**The joint committee process created student growth procedures favorable to teachers.**

Joint committees established local procedures for creating and approving the assessments used for student growth. Some student growth measures and scoring criteria resulted in low-rigor assessments—requiring the minimum amount of student growth, so that they were low-stress for teachers and permitted them to attain “proficient” or “excellent” student growth ratings with minimal effort. Vickie explained that the Abbidale school district joint committee was “very accommodating” of educators’ choices of assessments to utilize, while intending for the evaluators to encourage educators to increase the rigor of their student assessments:

Everybody had their own agenda in terms of student growth and what they wanted to count as student growth . . . it was very accommodating. It was accommodating on the teachers’ ends. It was what they wanted and the administrators were to agree with. During the evaluation process it was in agreement that it was still the teacher brought their idea to the administrator and the administrator would help them tweak it and try to make it a little more rigorous, if possible.
Glen explained that educators did not understand what rigor means: “There is a lot of confusion and there is a lot of debate about what constitutes a rigorous enough evaluation.” He often sent back assessments for teacher revisions “because they’re simply not rigorous enough.” In contrast, Rebecca felt her preschool educators created rigorous student growth assessments. She described her educators’ assessments as “developmentally age appropriate, but . . . rigorous.”

According to Aidan, his district’s joint committee set low targets to encourage teacher risk-taking. However, educators did not take risks; instead they adopted less rigorous goals because they feared a negative result on their summative evaluations. Aidan explained,

I don’t think we are at the point where it’s very rigorous. The teacher evaluation committee established through PERA had agreed to set the bar, what I think is relatively low in terms of the excellent and proficient percentages. . . . The thinking for the teacher evaluation committee was—if you set the bar too high then people aren’t going to be willing to take risks and really try something different and tinker with their assessments, because they’ll feel like, “Well, geez, why am I gonna set a real rigorous goal with the kids and set a high standard when obviously it’s going to put me at needs improvement?”

Jeffrey shared his observations of educators and student growth:

I have some teachers who I will say their SLOs have been extremely rigorous, and they haven’t been playing the numbers game of trying to get an excellent, but I have some teachers who’ve really looked at that closely and have tried to do that. I think, if anything, teachers have become a lot more savvy with data. In some cases, it’s a little too savvy and that they’re trying to game it.

Teacher-created and teacher-graded tests challenged participants. Corey and Charlie explained their districts’ decisions to use Type II and Type III assessments—tests selected, graded, and/or created by educators, as opposed to standardized Type I assessments created and scored by an external source —made it easier for their educators to manipulate the system. Charlie explained: “It’s trying to create ownership so that the teachers are owning it and doing it
themselves, but . . . it’s hard to police. It’s hard to monitor. That’s I think the biggest fallacy of the entire structure.”

Ann perceived that educators from another building in her district were “gaming” the procedures by “teaching to the test.” Harold shared an example from his previous district, noting how teachers used the student growth procedures to their benefit:

Yes, there are a number of teachers in my experience, primarily from my previous District, where their Type III assessments were vocabulary tests. They would give their students a vocabulary pre-test, they would teach a unit, give them a post-test on the same vocabulary and obviously those kids would grow significantly on that measure and they would say, “Here’s proof that I’m teaching my students. I’m an excellent teacher.”

Judging the quality of student growth assessments was challenging for joint committees and evaluators. Some decision makers lacked sufficient knowledge of the content or developmental level of the students being tested. For example, Charlie explained the difficulty of judging assessments when joint committee members were not specialists that content area:

Even though it’s gone through PERA [the joint committee], not everybody is a science teacher. Not everybody understands the sequence of math, and is that really at the right level of rigor? But then, you throw in all the other [subjects], like Spanish and art and music and food and consumer and P.E. Where is the commonality then?

Vickie also described the difficulty of understanding of assessment quality across different grades and developmental levels. She shared her experience of fifth grade teachers who were comparing their assessments to language goals set for kindergarten, English-language learners:

I’m in a K-3 building, so what a kindergarten teacher thinks is rigorous and a 5th grade teacher thinks is rigorous are two different things. When we’re writing SLOs, you have your kindergarten teacher who wants to increase the number of sight words. We have a baseline and they need 10 sight words in the beginning of the year and by February they’re going to know this amount of sight words, and trying to explain to them that that is not rigorous is hard because where is rigor in kindergarten when it comes to reading, and when you have kids that are not native Spanish speakers to begin with. When you’re talking comprehension and writing, that’s just a tough skill for our kids to begin with based on our poverty and where they have come. It takes a long time to get them to that, being able to put rigor into their curriculum. It’s been tough.
Procedures for student growth scores and the summative rating can inflate summative ratings. Joint committees established local procedures for determining the student growth scores used in the summative evaluation rating. These decisions included procedures for how student growth scores would be calculated and how the student growth and classroom observation ratings would be combined to calculate a final summative rating. State rules provided some direction for the joint committee’s decisions. In the first year of student growth implementation, joint committees could assign no lower than 25% weighting to the student growth scores, increasing to a minimum of 30% in the second year of student growth implementation. Joint committees who chose the default state model were required to weight student growth scores at 50% (ISBE, 2015). Among participants, 15 described problems with the procedures for student growth measures, including the growth targets set by educators, the procedures for calculating the summative evaluation, and the exclusion rules set by joint committees.

In many districts, educators typically earn the highest score on their student growth component. In her first year in her current district, Jordyn observed “probably 100% of the teachers always get four—‘excellent’—for student growth.” As a result of this experience, Jordyn set a goal in her second year to change the student growth assessments and goals to require more effort to earn an “excellent” or “proficient” rating. One educator who served on the Napa River joint committee admitted it was very easy to earn a high student growth score, as paraphrased by Ann: “The way this process is, if a teacher isn’t bright enough to figure out a way to show growth through this process, then they’re not really bright enough to be teaching anybody!”

Participants discussed two factors related to student growth scoring procedures. First, educators set student growth goals based on scoring targets and how many students will achieve
them. These procedures were established by the joint committee. Glen described the joint committee’s mindset when planning for student growth targets. While rigorous targets were intended, he described how an educator who sets challenging growth targets could be penalized:

We use an SLO (Student Learning Objective) model in my district, and I think the initial thinking, as to why the PERA joint committee wanted to go with an SLO model was, they thought it would foster more in-depth conversations around student assessments, how we analyze students’ assessments, and how teachers use those assessments to improve their instruction. All of which would be great practice, but I think tying that student growth model to their evaluation, actually puts those certified teachers in harm’s way when they try to create a rigorous SLO, because if they create one that’s too rigorous, that may be negatively reflected on their evaluation.

This potential penalty could discourage educators from setting higher targets. Instead, educators could set safer targets to lower their risk of earning a low student growth score. Jeffrey described how setting the targets was “like a game of just trying to meet the goal.” He felt sometimes he—as the evaluator—was to blame for not pushing the educators hard enough to set higher targets. Jeffrey also worked with educators who “worked really intentionally, and they’re really worthy goals. They’ve really helped to improve the learning of these students.” Glen questioned how educators could balance the mutual interests of setting challenging targets for learning while earning a high student growth score for their summative evaluation rating:

If I’m a teacher whose overall summative is . . . tied to how students perform on that particular assessment, I think the question becomes, “How rigorous do they really want to make it?” Can I say that I definitely got evidence of teachers gaming the system? No. Have I had conversations with teachers that have explicitly said, “Not only do I expect, but I want a 100% of my students to meet this growth target.” Yes, I have had teachers say that. And I don’t think that was really the intent of the legislation.

Second, participants explained how high student growth ratings—when combined with a low classroom observation rating—often resulted in a “satisfactory” summative rating. By their second year of implementation, most districts implemented the state minimum weighting for their summative evaluations: 70% for the classroom observation rating and 30% for the student
growth scores (ISBE, 2015). However, the procedures for combining these scores was a local decision, and many participants described how these procedures created “ratings inflation” that inhibited the assignment of underperforming ratings. Vickie summarized her frustration that student growth scores were driving the summative ratings: “If they get a high percentage on their student growth component, they won’t be in the ‘unsatisfactory’ or the ‘needs improvement’ category.” Charlie concurred:

When 30% of scores are already going to be proficient or excellent—for some teachers that are performing at a proficient level or maybe a little bit less—it’s hard to get anything less than proficient (on the summative rating). So, the system is kind of setup, in a sense, to create a large pool of “proficient,” hardly any “needs improvement,” and a few “excellent.”

Harold agreed underperforming educators would likely attain a higher summative evaluation rating in his district, overriding lower classroom observation scores from the evaluator:

Overall, the final evaluation or the final determination of the category that they fall into now with the student growth segment of it that is really left up to the teachers to control, they can essentially give themselves an excellent rating, with then offsets everything else that has been observed the administration. They end up becoming proficient instead of needs improvement or unsatisfactory.

As a result of summative ratings inflation, evaluators were hindered in identifying underperforming educators. For example, Charlie described his frustrations with summative ratings which prevented the use of improvement levers created by PERA and SB7:

It’s a little bit more of a challenge to get someone into a PDP because they would have to be pretty significantly low on Danielson, and I don’t see anyone really being classified that low on Danielson in order to qualify for a PDP. And so, to me it’s like I’ve got at least two people right now that I’m thinking of that I should put them on a PDP, but one day when you do the (summative) evaluation, they’re proficient. They might get a basic (“needs improvement”) here and there, but based on their evaluation, they end up squaring up proficient, and as long as that’s happening, they’re not going to get on a PDP.
Corey felt it was easier to identify underperforming educators before the policy reforms. In Ferndale school district, the summative rating procedures caused this difficulty:

By the time you have averaged domains one and four, which are not very dependent on the actual performance in a classroom, and you have thrown in what is almost always—at least—proficient student growth . . . in our district it is thirty percent of the overall year. . . It’s like everybody comes out at proficient. By the time you average everything, even if you are underperforming in domains two and three—by the time you average everything else—you are usually coming out at proficient.

So, yeah that’s a long way to get to, I feel like it was easier for me to, if a teacher was under-performing, to kind of use the evaluation tool and go at a specific area on the evaluation tool and not have to worry about all these other averages that are coming up.

Joint committees created procedures to determine which students could be excluded from student growth calculations. Student characteristics such as special education, Title I services, and English-language learners, as well as attendance and the number of classes a student must attend, were factors considered by joint committees (ISBE, 2015). One participant—Corey—described how the joint committee’s decisions in his district had the unintended consequence of reducing accountability, allowing educators to “cherry-pick” students:

We have allowed on attendance . . . you can drop the student’s score if they were below 80% attendance. But, what happens is you don’t have to, so you keep every kid who is proficient or excellent on your growth and then you drop the kids who weren’t. So, you pick and choose which kids below 80% attendance get to count [in your scores]. By the time all that stuff has factored and nobody is coming out below proficient on student growth.

The exclusion rules were meant to protect educators’ growth scores that could be negatively affected by students who were absent for a significant portion of instruction. Ann displayed empathy for a neighboring district with a high mobility rate, explaining noting how mobility influences student growth measures:

They have schools that have 100% mobility pretty much every single year. Imagine starting with 25 kids in August and you have a different 25 kids in May and you’re going to be evaluated based on the growth of some pool of those kids. The whole process, I’ve had issues and concerns with.
In contrast, five participants asserted the procedures in their districts did not negatively affect their ability to identify underperforming educators. For example, Dustin explained the decision grid in his district was adjusted after the district experienced student growth implementation. In identifying underperforming teachers, Glen explained, “I think it’s always been challenging, not necessarily just due to SB7 or PERA.” Molly stated the importance of documentation didn’t change under the reforms, stating “if you have something that is going to cause that person to get a ‘needs improvement’ in whatever area it is, you have to have that documented.”

**Evaluators’ workload and focus challenges their efforts to identify underperforming teachers.** Decisions about procedures influenced the roles of school leaders while placing an additional workload on evaluators. The reforms required evaluators to spend more time in classrooms for formal and informal observations. Participants explained the benefit of greater evaluator awareness of classroom teaching practices. Jeffrey described how his focus changed following the reforms:

> I think it’s changed my role to some extent. Not that I wasn’t focused on evaluation before, but it just seems like it’s a lot more intentional than it was in years past. In some ways, it felt kind of like a to-do; I got to get this done and do it and it’s done, but now it feels like it has a lot more intentionality to it. . . . I feel like I know my teachers’ practice better from doing evaluations in this manner.

Stephen conducted additional informal observations that yielded important information about teaching and learning in the classroom. Because the formal observation is planned and pre-scheduled with the educator, an evaluator “might see something in the informal that you might not regularly see nor necessarily see in the formal observation.” Stephen felt so strongly about the value of unscheduled, informal observations that he preferred to conduct informal
observations prior to the first formal observation. He stated he could identify more problems during the informal observation that could be addressed prior to the first formal observation.

However, the new evaluation plans required much more time for evaluators to complete the evaluation process. Eleven participants described problems with the increased workload. For example, Amanda described how the reform mandates impeded her ability to work with educators:

They’re taking a lot more time, there’s a lot more paperwork. I believe the paperwork is cutting down on the actual quality and time that principals and teachers meet to discuss the actual performance and ways to improve, everyone seems to be just completing paperwork.

Charlie elaborated on this point, focusing on the increase in face-to-face meetings required by his district’s evaluation plan:

I have over 150 meetings in one year to evaluate approximately 14 teachers. And, when you really look at the amount of time being spent on the evaluation, that’s a lot of time being taken away from kids and staff and walking around the classroom. I still try and get out at least to visit three or four teachers a day, as much as I can, whenever I can. But, that gets harder and harder to do when you gotta have three or four meetings a day for evaluation. . . . This can be streamlined. And, at the end of the day, teachers that are I know are proficient are still proficient. And teachers that I know are excellent—because they’re just dynamic people and really do a great job—are still excellent. It doesn’t take 150 meetings to determine what I already know is pretty much what it is.

Eric managed the workload by increasing his worktime outside of regular school hours to avoid reducing the amount of time he spent in classrooms:

It just means a lot more time that I spend outside of the school day. We’re not willing to give up what we do in the classroom or in the building, so I’m not doing evaluations at school or during the school day. So, my workday has increased.

The increased documentation requirements also changed the workload for educators who spent more time producing evidence for their evaluator to document and prove their performance level on the framework. They gathered documentation and evidence to share with their
evaluator—especially in the Danielson Framework domains 1 and 4. Charlie explained how this additional work was detrimental to their classroom instruction:

If they’re meeting a dozen times, 14 times in six months to try and prove themselves as proficient or better, they’re spending more time on the paperwork and the process than they are actually improving instruction. And that’s just the nature of it. They’re gonna spend more time on trying to prove it as opposed to actually doing it.

**Decisions by evaluators.** Bridges and Groves (1999) described evaluation decisions as the judgments made by evaluators in rating teachers and the actions that result from those ratings. In this section, the subthemes regarding interrater reliability, deferral of low ratings, unwillingness to issue low ratings, conversations about teaching and learning, Reduction in Force causes, the intermediate lever of PDPs, and remediation plans are explored.

**Interrater reliability is important, and follow-up analysis and training is needed.** When Illinois public school administrators completed training modules to become approved evaluators, the content focused on interrater reliability to improve the consistency of evaluation ratings statewide. Although the Danielson Framework was not mandated by ISBE, but it was recommended by education officials, and the statewide evaluator training utilized this framework and trained evaluators to achieve a high interrater reliability of evaluation ratings (ISBE, 2015). Nineteen of 20 participants worked in districts that adopted the Danielson framework.

Amanda stated interrater reliability is more important now because “this new system puts more pressure on us as evaluators . . . just as the teachers are defensive of what they received as a score . . . we have to more or less prove to teachers.” Following the initial training, some school districts provided additional training to ensure interrater reliability. For example, Aidan’s administrative colleagues watched videos together and discussed whether they were “in agreement in terms of what we would constitute as needs improvement, whether it’s in one specific domain on the rubric or overall.” After the initial training, Emma continued to work with
her two assistant principals on interrater reliability. She led shared walk-throughs, including conversations to debrief about the observations. Eric shared the benefits of “area colleagues having those communications and conversations about best practices.” This past summer, Vickie’s school district conducted interrater reliability exercises with all district evaluators. The next school year, her school district provided similar professional development for educators in the Danielson Framework, building staff capacity for interpreting the framework. Diana described how her district has focused on interrater reliability for evaluators and educators:

I feel that we’ve worked very hard to make sure that administrators have the support as well as the teachers. . . . We have several committees that work on interrater reliability and we look at our PERA practices every year, and changing what needs to be updated and tweaking. We listen to our teachers who are part of the union as well as social workers.

Four participants cited problems with interrater reliability within their districts. First, the procedures were not consistently followed. Amanda believed principals did not follow the procedures of the evaluation tool, citing “complicated systems and processes” and the constraints of evaluators’ workloads. Second, low interrater reliability resulted in uneven distribution of summative evaluation ratings across schools in the same district. Emma observed discrepancies in ratings distribution, explained her superintendent discovered more elementary educators rated as “excellent” than at the middle school. Emma attributed this discrepancy to evaluators’ individual interpretations of the Danielson Framework. Corey described one meeting when the distribution of summative ratings from all the buildings in his district was shared, revealing problems with interrater reliability cross the district. Eric explained that his colleagues have not discussed interrater reliability for about five years. He feels it has become “more of a personal concern that it’s done with some fidelity than truly checks and balances on our end.” However,
because evaluators are now required to complete mandated retraining by ISBE, Eric anticipated an increased focus on interrater reliability in their districts.

Participants were asked to estimate what percentage of educators in their school and in their district were underperforming. Their estimates ranged from 0% to 20% (Table 16). Further analysis of this data will occur in the integrated findings section of this chapter.

Table 16

*Interview Participants’ Estimates of Underperformance*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Estimate of underperforming educators in participant’s school (% of underperforming educators)</th>
<th>Estimate of underperforming educators in participant’s district (% of underperforming educators)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aidan O’Brien</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Amanda Ashbee</td>
<td>&lt; 2%</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>Ann Keaton</td>
<td>0%</td>
<td>5-10%</td>
</tr>
<tr>
<td>Charlie Ramirez</td>
<td>10-20%</td>
<td>10-20%</td>
</tr>
<tr>
<td>Corey Hammer</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Diana Legend</td>
<td>4%</td>
<td>8-10%</td>
</tr>
<tr>
<td>Dustin Ross</td>
<td>5%</td>
<td>15%</td>
</tr>
<tr>
<td>Emma Moore</td>
<td>10%</td>
<td>15%</td>
</tr>
<tr>
<td>Eric Graves</td>
<td>&lt; 5%</td>
<td>&lt; 5% (*)</td>
</tr>
<tr>
<td>Frank Steele</td>
<td>1-2%</td>
<td>1-2%</td>
</tr>
<tr>
<td>Glen Tucker</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Harold Cooper</td>
<td>&lt; 2%</td>
<td>5%</td>
</tr>
<tr>
<td>Jeffrey Shannon</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Jordyn Harris</td>
<td>20%</td>
<td>No opinion</td>
</tr>
<tr>
<td>Melody Roberts</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>Michelle Hale</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>Molly Martin</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Rebecca Silver</td>
<td>5%</td>
<td>15-20%</td>
</tr>
<tr>
<td>Stephen Pierce</td>
<td>5-10%</td>
<td>5-10%</td>
</tr>
<tr>
<td>Vickie Wells</td>
<td>2%</td>
<td>2%</td>
</tr>
</tbody>
</table>

*Low ratings are deferred due to retirement and/or seniority.* Five participants deferred assigning a low summative evaluation rating due to the potential for the educator to retire and/or their seniority in the district. Aidan explained how he weighed various factors when deciding to not to assign a low rating for a senior educator:

I think there are times where you feel like people are not up to snuff and you have those conversations. I think where I’ve been guilty of backing off a bit is it’s someone’s maybe it’s their last few years, they’ve already put in their retirement papers and you’re kind of
like, “Is it really worth it?” That probably sounds terrible, but I think it’s just trying to be respectful of where they are in their career and how do you coach ’em up a little bit and navigate that rather than, “Now we’re going to go after you here and you’ve got two and a half years left before you retire.”

Corey used a practical lens when considering an underperformance plan. He explained, “When you are 3 years from retirement, by the time the (improvement) process would play out, you would be retired anyways.”

Amanda described how seniority influenced her decision to decline issuing a “needs improvement” rating to a senior educator. In order to justify a higher rating, Amanda included language in the narrative evaluation that explained “needs improvement” performance but only scored a few specific components as “needs improvement” so the overall rating would remain “proficient:"

I’ll be 100% honest and I will say, yes. If I’m here with a tenured teacher who’s been in our district 25 plus years. . . . This evaluation system is new . . . do I put a proficient or do I put a needs improvement on this teacher, who in the past before me evaluating them got nothing but excellent, nothing? Now, here I step in and I’m going to tell this teacher they need “needs improvement?” So, once again what will I do? I will put the proficient down but my statements will be statements of . . . needs improvement. . . . So, you have to be careful. It might do you more harm than good by putting a needs improvement in this category and just a proficient. Might not affect the overall rating at end but that teacher is going to see that needs improvement and that’s going to be like a bullet, and then you’re going to have the ripple effect. So, I’ll be honest: Yes, I’ve struggled with that.

Michael evaluated an educator in her 43rd year of service. Although the educator’s instructional practices reflected underperformance, Michael deferred the low rating out of respect for her seniority and potential retirement plans:

I knew she had already had some conversations with me about retirement; I knew that was on the horizon. I thought after 43 years she deserved the opportunity to go out on her own. She wasn’t doing any harm to kids. It was again that instructional practice. She knew her stuff, she knew her craft, she was still able to teach kids. Was she student centered? No, but the community embraced her, the kids embraced her, she still had strong relationships. I gave her a pass. If I was going strictly by the book, she would have been on a remediation plan.
Charlie described a case when he negotiated with an underperforming educator over her reassignment to a new grade level. He used the reassignment as leverage to obtain an earlier retirement by one year:

Had we pushed a little bit further, probably could have continued on to get her out. But, she was close to retirement so on her second to last year, basically I went up to her and I said, “We’re still not changing, even though we went through a PDP process,” and eventually she did enough to get off of the cycle. The instruction is still not there. One of the things I told her was I was going to move her to a different grade, and she made a deal with me, in that if I could give her one more (year), she would retire at the end of the year, and I held her up to it and she retired.

In contrast, Eric explained the process must be honored, noting that declining to issue a low rating would be unfair to other educators in the building. In addition, Ann has not avoided assigning low ratings when they were warranted. She explained, “Just because they’re retiring doesn’t mean that I would give them a free pass.”

**Evaluators are sometimes unwilling or unable to issue a low rating.** Nine participants described situations when they were unable or unwilling to issue a low rating to an underperforming educator. First, when school leaders build relationships with staff, it can be difficult to separate personal feelings from personnel decisions. Charlie described the struggle evaluators faced when making difficult decisions about staff underperformance:

Right now, I’m in a tough one because I have a teacher that’s very, very well liked. I think she’s very charismatic, very caring and involved, but her classroom performance. . . . I mean, barely proficient. I mean, it’s just proficient in third year and I think I’m going to have to let that teacher go. It’s not anything to do with PDP or anything, but that’s where I struggle. I struggle with teachers when they’re good people. They’re hard working people. They just don’t have the skillset to bring it up to the level we need in order to get our kids where they need to go, and it’s really tough because it’s a very amiable person. That’s where I struggle, because I want to see this person be successful, but after three years, I’ve got to make a call.

Second, evaluators may issue a higher summative rating to avoid conflicts, either with an individual educator or within the overall building, which could affect the school culture and/or
good will within the community. Corey described a popular educator in his school who “does all the intangible things that you need.” However, Corey feels her instruction is problematic. He feels using improvement levers from the evaluation process could backfire, as he explained she is “not a safe person to do that with . . . because she has a lot of respect in the school and a lot of credibility beyond the school.” In this case, he preferred to use improvement levers outside the evaluation process to avoid the political problems and potential negative effects on the building culture it would create.

Third, evaluators may adjust their ratings to reflect the summative ratings inflation in their district. Corey cited an example of a new educator whose performance reflected the “needs improvement” descriptors. However, due to summative ratings inflation in his district, very few educators ever were identified with “needs improvement” language, and he felt assigning the accurate rating was unfair when comparing her performance to summative evaluation ratings issued to other educators in the district. Corey concluded, “To get proficient should be hard, and it’s not anymore.”

Finally, evaluators may be influenced to retain an underperforming educator for reasons other than their teaching performance. As a new principal, Jeffrey was sensitive about diversity on his staff. Midway through his first year, he recommended rehiring an underperforming educator for a second-year contract—based upon the person’s ethnicity from an underrepresented demographic group:

This goes back a few years ago, and at the time, and yeah, I’m not real proud of it, but I did change my view on the whole thing. I was in a building that had no African-American presence whatsoever. Our district is always looking at trying to increase the number of African-American teachers in our buildings, and so I was assigned someone by HR. I got this guy, and I’d known him for quite some time . . . he was a good guy, but he was just not a good teacher.
I think my initial evaluation the first year that he was there, I was probably kinder than I would have been otherwise. I did assign him a mentor teacher, but as far as a reflection on his evaluation, I would say it was probably not as reflective of his performance. I wasn’t real happy about making that choice, but I think in the second year that he was in my building, I just could see it wasn’t going to be a good fit. Ultimately, teaching was not his thing, so I ended up not renewing his contract the second year.

**Conversations about teaching and learning have improved under reforms.** Evaluation reforms increased the number of conferences between educators and evaluators. With the aligned focus on improving teaching and learning, evaluators increased their need to hold difficult conversations with staff members about their performance. Nine participants described how conversations about learning, student data, and instructional performance improved following the reforms. Amanda explained how these meaningful conversations were more important than the evaluation process itself:

> I do not think that the system itself has improved instruction or teaching, I believe it has to do with the evaluator and the administrator who offers the support and has those meaningful conversations. I don’t think a system or a documentation improves instruction. My teachers have always been open and willing to improve instruction. They do not like the new evaluation system any more than I do, but we both take it for what it is, and we both complete it the best we can and we do it meaningfully.

Michael found his professional conversations were “much deeper, both about their practice and their individual students.” He explained educators shared data with their evaluators and discussed performance of individual students. He described this dialogue as “transformative . . . causing shifts in individual practice.” Diana felt the framework provided objective criteria for professional practice that depersonalized difficult conversations about underperformance: “I really think just having a research-based place to start helped with that communication.” Her district revised the Danielson framework to create a local framework for early childhood education, because many educators felt the framework had limited application to developmental levels and practices of early childhood education. A local committee was formed to create the
new framework that included early childhood teachers, administrators, and union leadership. She credited this work with reducing stress on educators, “because it’s clearer to teachers. They don’t feel they have to guess what to do to be where they are on the rubric.”

In contrast, Amanda felt educators are too focused on their summative ratings. When meeting with their evaluators, they might minimize the importance of their professional conversations and the opportunity to obtain feedback—choosing to primarily focus on the summative rating instead:

To me it’s horrible, because my teachers in the past, they can have a conversation about what has gone well and what needs to be improved upon. But it’s a stupid rating of either “excellent” or “proficient.” Excellent and proficient—they get hung up on the word and not what led to that rating. All they look at is that word. Whatever that score is . . . the final word. I have teachers that turn right to that back page and just want to know what the final score is and what word they got, whether it was “excellent” or “proficient.” And if they got “proficient” instead of “excellent,” or they got “needs improvement” on a category (component) instead of just “excellent,” their whole world falls apart!

Michelle agreed that professional conversations are difficult when educators realize they have received an underperforming rating:

As soon as you give somebody an “unsatisfactory,” they’re done. That’s like the anti-Holy Grail in the Austin district. If you issue an “unsatisfactory” in any component, the person is looking for a job the next day, which is probably fine, because it’s pretty hard to get “unsatisfactory.” That then deteriorates the ability to coach.

**Reduction in Force was seldom used.** The state’s Reduction in Force (RIF) reform was a seldom-used lever to remove underperforming educators. Five participants discussed circumstances that lead to a RIF action in their district. District finances were one reason educators are released—positions were eliminated due to funding. Program changes were the second reason—changing needs in the school necessitated the employment of educators with different certification. The third reason was declining enrollment that necessitated staff reductions.
Dustin’s school district experienced two RIF actions due to the economic recession:

The recession hit our county quicker and harder than others, so we had a decrease of enrollment because of that. We also had a tremendous amount of people who were at the very top of the salary schedule, so we had some expenses that were not controllable.

The economic and enrollment situation was known to his staff ahead of the RIF action, and Dustin explained how staff members were discomforted by the effects of the unknown outcomes:

I just think people were always tense and wanting to make sure that they were in the necessary group as far as evaluation ratings go. Then they also were conscious of how many cuts did we have to make, and where would we make those.

Michelle’s school also experienced financial challenges that resulted in staff reductions:

The Austin district, over the last seven or eight years, has had some real trouble with finances and things like that, so I’ve had . . . three teachers who have been laid off due to budgetary reasons. . . . In the Austin district, [educators] who are rated the same are then listed by seniority . . . last-in/first-out. So, anybody who’s in the “basic” [“needs improvement”] summative rating, they’re all looked at together and then the proficient people are all looked at together, the excellent people are all looked at together. The unsatisfactory people are the first to go. . . . Two of the three . . . it was crushing to reduce force.

Due to programming changes Corey experienced a RIF situation at Ferndale school. His tenured educator had been identified as underperforming, but the district modified her position.

Later, the position was eliminated, resulting in her permanent release:

I had a teacher who . . . went on a Professional Development Plan, and then we were going to do the remediation plan. A change of position was then made. She was assigned to a different position in the district—in the same building. It was a unique position, it was only for her, and she would then have fallen into group two at that point. She was continuing to underperform in this new role, and then we got to a point where we decided to eliminate the position. Because she was in group two, we eliminated the position, which caused her to be RIF’d. . . . At that time . . . members of the union in group two did not have recall rights. So, she was eliminated. Sandy (pseudonym) was the teacher that we released through the RIF. She was on our remediation plan and Professional Development Plan and multiples over the years and then it was the RIF that finally was able to . . . help us to get her finally out. But honestly that’s the only employee I have done this with, where the goal was to get them out.
Changing enrollment resulted in a RIF at Michael’s school. A non-tenured educator was in a lower evaluation tier because of her “needs improvement” summative evaluation. Consistent with the new state policy, she was released instead of other educators with higher evaluation scores. Diana experienced two RIF actions; both educators had been issued “needs improvement” summative ratings and were non-tenured.

*Professional Development Plans were an intermediate improvement lever—yielding mixed results.* Eleven participants created a PDP for one or more educators. Corey described PDPs as “a fair process . . . to give the teacher a chance to improve.” Participants shared cases of successful and unsuccessful PDPs they participated in.

Melody described two successful PDP cases she supervised. First, Melody described a 25-year, tenured educator who was “stuck in her ways of teaching and had a hard time adjusting to the . . . new challenges that we’re having with the students.” Melody integrated building-wide professional development into the goals of the PDP, including trauma-informed care training for the entire staff. She reported the plan was successful, as it gave the educator “the tools and strategies to be able to support the children that she has in her classroom now.” Second, Melody described a case involving a social worker who successfully completed his PDP. Melody described using professional development from internal and external sources to successfully improve the weaknesses, including time management and parent interaction. Rebecca also shared positive outcomes of a PDP that resulted in teaching improvements:

The teacher is now really spending time on lesson planning, including higher-level questions, materials . . . communicating with parents, attending professional development, teaching lessons, instead of just giving an open-ended, free-for-all, go-play-in-the-center-you-want-all-day. It’s become more learning-based, standards-based.

Other participants shared mixed results from PDPs. Michelle implemented a plan for an educator she described as her “worst performing teacher.” The plan was started, but the educator
attained a subsequent evaluation rating of “proficient”—meeting the expectations of the PDP—which allowed her to continue in her position, despite “still being a bad teacher.” Michelle explained the criteria for earning a “proficient” rating in her district was easy to meet—but the criteria did not reflect her own professional expectation for satisfactory teaching performance.

Glen’s experiences with PDPs resulted in one educator who improved but two educators who did not improve to the district’s minimal performance expectations.

The PDP could be used as an intermediate improvement lever when an educator needed to improve—but the inadequate performance was not deficient enough to warrant an unsatisfactory rating and placement on a remediation plan. Michelle described how an educator with an unsuccessful PDP experience transitioned to a remediation plan in her district’s evaluation process. Jeffery illustrated how he used a sequence of steps, including developing a PDP:

I issued him a Notice to Remedy, and the following year we entered into a Professional Development Plan. We set some targets. Most of them were related to classroom management and his ability to just follow through with grading and planning and those kinds of things. We set those targets. He had a mentor teacher, or actually, no, I take that back. He didn’t during the Professional Development Plan. That’s not part of our district plan when they have that Professional Development Plan, but certainly we offered support within the building. He failed to meet those goals, so then we entered into a remediation plan where he had 90 days to make changes in all of the identified areas.

Charlie implemented two PDPs for educators. His first educator successfully completed the plan. Charlie suggested the plan “made a big difference and made a great improvement.” In Charlie’s second plan, he felt the process had a positive but “minimal impact,” because the educator learned they could “only go so far and not push anything lower than that.” Because this second educator continued to struggle, Charlie proposed moving her to a new grade level. Instead, she asked to remain at her current grade level in exchange for retiring in one year, to which the district agreed.
Remediation plans usually result in the educator’s separation from district employment. Among participants, 12 had implemented one or more remediation plans. In describing their experiences, only three participants described cases that resulted in the educator’s improvement and continued employment. The others ended in resignation, a negotiated settlement, or dismissal. In Ann’s district, the board and educator negotiated a separation agreement that included a lifetime agreement to not seek teaching employment in Illinois. An educator in Corey’s school retired upon learning a remediation plan was imminent. Molly’s employee resigned with 20 days remaining on the 90-day remediation plan. In Melody’s school, the teacher was allowed to voluntarily resign prior to board action for dismissal. In contrast, Dustin described how remediation helped his teacher turn around their classroom management skills. The district provided models and intense professional development that facilitated sufficient improvements in classroom performance.

Actors and their access, interests, and power. Bridges and Groves (1999) defined actors as persons who “figure prominently in the politics of personnel evaluation” (p. 323). The following subthemes present data on the influence of teachers’ unions and their members, school board members, superintendents and central office personnel, parents and students, and educators and the overall building climate.

Union influence: Balancing employee rights with a mutual interest in effective teaching. Eleven participants described neutral to positive influence from unions when educators were identified as underperforming. Union representatives followed the procedures of the evaluation plan but shared a mutual goal shared with the school leadership: the desire to support quality teaching the classroom. In many cases, union representatives were helpful and
supportive. According to Diana, the union representative shared the district’s vision of educator quality:

I got the feeling that the union representative clearly saw the weakness in the teacher and we have heard in our joint committees that the union’s goal is also to not have weak teachers in our district because it is such a drain on the district and, so not a good idea to keep supporting bad teachers. . . . I felt very comfortable having the union president sit there because I’ve been in meetings with them, so we’ve already talked about, “This is what we’re going to be looking for. This is what we would like to see in our classrooms.”

Glen described how his union acknowledged the need for underperforming educators to improve “for the betterment of the entire building.” Emma also shared how the union supported her efforts to improve instruction:

I think they understood that at the heart of the plan we were trying to improve instruction and make sure that the individual was maintaining appropriate professional relationships. And so, I think on behalf of the union as a collective whole, they wanted what they saw was best for like a safe and healthy building environment. So . . . my perception was that they were in support of it.

Jeffrey felt the union reinforced his efforts to address underperformance, including counseling the educator to consider a different profession.

During a PDP experience, Charlie worked closely with a union representative who attended most meetings with him and his underperforming educator. During these meetings, discussions included reviewing and updating the plan, debriefing about observed lessons, and delineating the needed changes. Charlie described how this representative engaged the educator in tough conversations about practice, assuming the role of an instructional coach:

To be honest with you, the guy that’s still here in my school today, the reason he’s still here today is because his union representative basically told him to shape up or ship out. And, she was done. She basically told him—which I really appreciate her. . . . She was mentoring him, and she basically told him “Dude, either do this and do it right, or I’m out of here and you’re on your own.” And so, she basically told him he needed to shape up, that everything I was putting up, it was the truth—that I wasn’t exaggerating anything, it needed to change. Because she was pushing him, she was basically mentoring him. That was a great way of leadership from union representative coming through and coaching another teacher. It’s about as close to a coaching model as you can get without a coach.
But, she was very knowledgeable, and she basically told him how to change, and he did. But, if it wasn’t for her, I think he probably wouldn’t have made it on his own.

Positive union relationships are important to maintaining the building climate. Diana welcomed union involvement, because she “would prefer the union president to be there to support a teacher who’s failing in my opinion, underperforming... so that she can get the help and be a better teacher.” Jeffrey believes his relationship with the union was strengthened by approaching the underperforming educator with a respectful demeanor:

My union president actually was very complimentary to me and actually shared that with the superintendent just because, I don’t know, I didn’t get pissed off. I think that can be the case. I treated him with dignity and respect, and I think that’s a key. At the end of the day, he was a great guy. I probably would’ve gone for a beer with him, but he stunk as a teacher. I just was very respectful in how I dealt with it... You can be an asshole, and yeah, that would give you the same results, but you might as well just be respectful. It’s frustrating as all get up, don’t get me wrong, but I just continued to be respectful. That’s my personal philosophy. I just try to be respectful of whatever position you have in my building. If you’re not performing, I’m still going to be respectful, but I’m still going to expect that you’re going to continue to do your job.

When union advocates or representatives were involved, these individuals were impartial to positive in their influence. Diana described a neutral relationship with the union representative:

He never said anything to me. We never went there. He and I kept it very professional and neutral. Now what he said to the teacher when I wasn’t in the room, I have no idea. But I did not get any fallout behind the scenes or in any other way... I know he was emotionally supportive of her because it was a very emotional time. We went through lots of boxes of Kleenexes and he was very emotionally supportive.

Melody shared a similar impartiality with the union representative in her previous district—describing how the representative would “mediate the process:”

They do not try and sway administrators in either direction. They will just sit and go through the facts between the teacher and the administrator. I know they meet a lot with the teacher behind the scenes, and they work with them a lot... I think they’re a positive support when it comes to remediation plans or underperforming.

This previous district had a weaker union presence, according to Melody.
Michael described the importance of building positive and respectful relationships with the union: “I know what they’re doing and I know what their role is. I’ve always affirmed that and acknowledged that with them. That’s allowed us to grow as professionals and have deeper relationships and very respectful relationships.” Michael also shared the importance of union representatives having direct knowledge of the educator’s performance problems. When the observed teaching practices conflict with their professional values, the union and its representatives are likely to support efforts to improve underperforming educators:

They were in the building so they had a firsthand, first witness account to what was going on in terms of why that teacher needed to be remediated. It wasn’t coming to them through hearsay. I think that was something that was really in my favor or in the processes’ favor to have someone at the ground level to know that yes, what they’re doing and their instructional practices and how their speaking to children, is not right . . .

I’ve been fortunate to work with unions that held pretty student centered beliefs in their philosophies and their approach to education. I think that went a very, very long way, the fact that they were able to see that. They saw that one of their own members were in violation. It’s not to say that they’re going to go out publicly and say that but they knew that member was in violation of what their own values were as an organization. They knew it firsthand. And they knew that it wasn’t because somebody got into a disagreement or this evaluator was doing this to be punitive or anything. This really was going against their own, as I said, values and norms.

Molly described working with the Illinois Education Association regional director to select a consulting teacher for the remediation plan. All parties agreed to find a special education teacher from outside the county to mentor and assist the struggling special educator. Despite the improvement efforts, the consulting teacher shared with Molly her frustrations about the lack of improvement shown by the underperforming educator:

I was having candid conversations with . . . the mentor. She was coming to me and she was actually opening those conversations. . . . One time, she said, “I don’t understand. I don’t know why she doesn’t just do what you’ve asked her to do. You haven’t asked her to do anything that is out of the ordinary.” She said, “This is what all of us do in special education. I don’t understand why she is fighting it so much. She should be able to do this, but she is not going to.” She told me about halfway through, she said she is not going to meet the remediation.
Ann also felt the union was supportive during a challenging remediation case. After 2 years of documentation and remediation efforts, both the state and national union representatives suggested the educator negotiate a separation agreement with the district and agree to leave the teaching profession:

We gathered evidence over a 2-year time and it took us then about seven months of legal battles. Then finally, the IEA—and actually NEA—finally told him, “Look bud, your best chance is to go into what’s called a separation agreement.” The district agreed to give him any accrued sick days and he had to agree to giving up tenure and never seeking employment in an Illinois school again.

Frank shared a case in which the union created friction over comments an evaluator included in the summative evaluation notes. Although the educator earned an “excellent” rating, the evaluator noted “specific behaviors would need to continue to be maintained for that person to continue on receiving an excellent rating in the future.” According to Frank,

The association had come in and demanded that that be removed. While it didn’t necessarily try to influence a higher rating on that one, there was some definite friction between what the association and we, as administrators felt to be included in evaluation and observation document. I think it had some long-term effects, specifically between the (evaluator) and that teacher. I think it continued a pattern of friction between the association and the district leadership. I think it was symptom of a larger problem.

School boards supported the evaluators—but navigating the political terrain helps.

None of the participants cited the school board as a hindrance to the identification of underperforming educators. Board members may have provided information but did not direct the evaluator on how to use the information. Michael noted board members may “provide insight but never once tried to influence me one way or the other . . . never once talked about performance ratings or anything.” Vickie concurred, explaining board members “think they always like to throw their two cents in but it’s not to persuade you to do anything. Certainly not with any leverage.” Glen shared how a board member provided information heard from parents or the community but did not attempt to influence decision making:
I have had a board member that was also a parent in my building. He had, obviously, first-hand knowledge of some of the teacher performance issues that I had in my building. I think he would have fairly candid conversations with me, about what he hears in the community, but at no point did he ever explicitly try to influence anything I was doing. I think maybe, implicitly, he just wanted to me know, “This is the word on the street. You’ve got your finger on the pulse of this.”

Though school boards may not interfere with the process, skillful political navigation is required when relationships exist between board members and underperforming educators.

Molly described a challenging case that required political navigation of community relationships and board politics. She expressed unwavering confidence in her independence as an evaluator, explaining “no one has ever tried to influence me because they all know me . . . they are not going to influence me.” Her resolve was tested, however. The underperforming educator was friends with a board of education member; consequently, the board took an active role:

It was hard to let (the teacher) go because she had a friend on the board, and this friend on the board just really did not want to let her go but (the teacher) was a mouthpiece and she collected information here, inside the building, and shared with that board member. Stuff (the teacher) wasn’t supposed to be sharing for sure, so that was . . . the tricky part of all of that.

Through this experience, she learned the necessity to keep the board informed about underperforming teachers, especially in a small or rural community, and having the full support of the board president:

After I talked to the attorney, I called the board president at that time, the same day, and explained to him what had happened. And he said, “Well, if that is what it is, that’s what it is, and we’ll do whatever we have to do.” And from that moment on, which the board president pretty much led the board meetings with me just making reports. We tried to keep it as hush-hush as possible, but it’s a small community.

The school board took an active role in Molly’s case due to the sensitivity and involvement of the community. For example, Molly was asked to present data to the school board and created spreadsheets illustrating how many students earned failing grades. During the process, the board asked to meet with the underperforming educator in closed session. However, the board
ultimately supported Molly when they released the non-tenured educator, with the educator’s friend on the board abstaining from the vote.

**Support from the superintendent and central office administration is essential.** In most cases, the superintendent and central office were supportive of evaluators’ efforts to improve educator effectiveness. Only one participant described the central office as a hindrance to the identification of underperforming educators. In larger districts with large central office staff, human resources staff provided the principals support as they developed and finalized their evaluations and improvement plans. When asked about central office support, Michelle responded,

Yes, 100%. I mean so much so that when a person is on a remediation plan, you get a call from central office and they make sure you know how to do the process and they’ll walk you through all the paperwork and everything. Like you don’t have to initiate that call, they’ll call you.

Charlie explained his human resources director often keeps track of educators on his evaluation “watch list”—those who demonstrate some level of performance concern to Charlie. These conversations with human resources are not meant to influence his evaluation decisions, but rather to keep the central office from being “blindsided” by a low rating. Charlie felt supported by his superintendent and central office:

They always come back to, “It’s your building. You’re running the show. You’re the one that’s going to be accountable for the scores and everything else happening in your building, so if you’re good with it, understand that it’s your call.” So, they pretty much put it on us to kind of decide how we want to shape our building within the district expectation. So no, I have not seen that from anybody for that matter.

Aidan and Stephen described similar support from the central office, such as sharing information perceived as informative—but not intended to directly influence the principals’ determination of summative evaluation ratings. Aidan explained,
People (central office) have conversations with me and say that “Wow, this teacher is really impressive in some of these district-level meetings,” or “Wow, they had a really bad attitude,” or “I’ve gone in their classroom and it was really poor.” Then you just have to follow up and get in the classroom yourself or have conversations.

Vickie described the vision of her superintendent regarding underperforming teachers. The superintendent supported several remediation plans and backed the evaluators’ decisions to address underperformance:

The superintendent I’ve always had in this district has always encouraged you to “call a spade a spade”—and if the person’s not doing what they need to be doing, to take the necessary steps to make them a better teacher. We’ve gone through quite a few remediation plans within our district. I can think of at least one at the junior high and another one from another related service person that was also shared that went to remediation at the junior high. I think the high school principal took another person through a remediation plan at the high school. We haven’t been told not to do it.

Vickie also shared an anecdote she has heard about superintendent support in general for remediation plans. She was frustrated that a central office leader might factor in the financial cost of remediation when determining whether to pursue an improvement plan:

I hear about superintendents who say, “We’re not going to do remediation plans because they’re just too expensive and time-consuming and costly for the district.” What good is SB7 and student growth if it’s a financial decision—in the end—if we’re going to take finances over what’s best for kids in the end?

In contrast, Glen experienced varying levels of support when he was first hired by Crater Bay school district. He identified three teachers whose performance warranted PDPs, but his central office prevented him from starting a plan for two educators. The superintendent and central office administrators cited concerns about potential disability claims that could be leveled against the school district due to the teachers’ age and/or disability status. Therefore, Glen did not identify these two educators for underperformance, based upon central office directives. However, in Glen’s fourth year, a new superintendent was hired who shared a firmer stance regarding addressing underperformance, and Glen started PDPs for the two educators:
Once the superintendent changed, it was clear by what he was communicating—not only to me individually but also to all of us as administrators at district leadership team meetings—that the whole tenure was going to change in our district. And so I think it was my fourth year then, that I put two more teachers on a Professional Development Plan, because I felt I now had the support to give them a rating that was more reflective of their performance.

Glen now felt supported by his central office. He explained, “I have reached out to our assistant superintendent of curriculum instruction, or personnel, to come in and do additional observations, or to provide additional feedback on overall work performance.” Jeffrey’s human resources department provided sample plans, while Harold received support for the timelines and procedures.

**Parents and students provide helpful information to evaluators.** Participants described the influence of parents and students as indirect influence—by providing information that evaluator might choose to more fully investigate. None of the participants described parents and students as a hindrance to the identification of underperforming educators. As Aidan explained, “The outside public, students and parents don’t know what teacher ratings are, but they are certainly not shy to share information, whether they think the teacher is doing a great job or they are not.” Glen speculated, “When you’re a principal and you become established, I think parents feel more freedom and more opportunity to share their candid thoughts on particular teachers. That does occur, but I don’t know if it’s intended to change ratings necessarily.” Jeffrey explained may investigate a parent complaint, but the complaint doesn’t directly influence his evaluation. Emma suggested that some evidence from parents may influence one evaluation component related to parent communication, but not the evaluation overall.

Several participants shared accounts of parents whose information provided data that was helpful to them as they addressed underperformance. For example, Corey described a case when parent information indirectly influenced his evaluation:
I had so many parent complaints before I even ever observed her, that I am positive that had an influence on what I saw in that classroom. But it really was that bad. . . . The parent complaints also supported me in being able to pull her out of that traditional classroom setting. . . . The district supported me moving her out of a regular teaching position at the semester because of the parent volume happening around, so it was helpful.

In another example, Ann described how parents shared information with her during a challenging case with an underperforming art teacher:

I certainly never had a parent tried to influence how I evaluated any of my teachers. . . . I had a lot of parent complaints—and valid complaints—that influenced her evaluation. The parents certainly didn’t ask for it to be a part of the evaluation. They were just calling more in kind of, “Hey, my 3rd grader used to love art and now every Wednesday he has a stomach ache and doesn’t want to go to school because he has art.” It influenced it from that standpoint, but I’ve never had any one directly tried to influence it.

Finally, Michelle described a case in which parent feedback provided valuable information that helped her persuade an educator to resign:

I had a teacher last year who was phenomenal at the dog and pony show, but terrible otherwise. So she actually rated out fairly well. She was a probationary teacher. She rated out fairly well, so when it came time in the spring to decide whether or not to renew or non-renew her, I didn’t have the opportunity because her ratings were good enough that she wasn’t even in range to be non-renewed. But we had so many parent complaints and so many other issues from her class, that I was able to convince her that teaching was not her calling, which was good, but I had lots of parents ask me why she was teaching in our building, and I think they were right.

Evaluators in rural areas of the state must be careful navigating the sensitive politics of a small, connected community. Vickie described a case with an underperforming educator when both parents and community members would share helpful information with her, fueled by the teacher’s poor reputation in the community:

We’re a very small town. Everybody in Abbidale, Illinois (pseudonym) knows everybody and people know who the good teachers are—according to the community—and people know who the bad teachers are—according to the community. You have those teachers that are requested and teachers that aren’t requested.

When you have a teacher that’s being inappropriate outside of school you know who that is. It’s interesting every now and then you’ll get a phone call from a parent that says, “Do
you know what so and so did this weekend?” It’s like, “If it’s not affecting their job performance, it’s really none of my business.” It does make you look at that teacher a little harder on some days and say, are the lesson plans written and you may not have done that had you not gotten that phone call.

Influence from parents or students is not limited to negative feedback. Corey described how positive student feedback influenced his opinions of educators:

If you are getting great feedback from students, does that change my view of a teacher? Very much so. If all I’m hearing from kids is that this teacher is mean and this teacher is not good, that influences how I see that teacher and I think it’s the same way from parents.

**Educators ultimately support improvement, but the school culture may suffer initially.**

Six participants reported the identification of underperforming teachers *ultimately* improved the overall culture in their schools once the problem had been remediated—either through teacher improvement or teacher removal. However, four of these participants reported the *initial* school climate became somewhat negative.

Rebecca explained the initiation of improvement plans affected school culture negatively because many teachers—especially veterans—had deep friendships with an underperforming colleague. She reflected, “If you put someone like that into the (improvement) program, you might not be popular.” According to Frank, the process of remediation caused initial unrest among his staff:

I would say during there is definitely some unrest among the staff specifically. I think there is also unrest when people recognize that things aren’t going well within a classroom, when they recognize that classroom management isn’t going well or they recognize those kinds planning issues are coming up within a classroom, I think there’s unrest regardless. So this was once that person had been informed that they would not be returning, I think there was some unrest among the staff, simply from a discomfort, not necessarily from an anger or hostility, but just a discomfort that there was someone who had been told that they would not be returning.

Educators often supported their colleagues who were identified as underperforming. Glen described how teachers formed coalitions to protect their peers: “There was always kind of a
‘circling of the wagons’ whenever someone within the association is getting the attention of the administration about their performance. I think that’s something that you just have to weather administratively.” Harold also described a “circling of the wagons” as teachers rallied to support their colleague:

That’s where we struggle, it’s the unspoken issue in school administration. As soon as you start identifying teachers for remediation, or you’re identifying that they’re not as proficient as they think they are. Teachers “circle the wagon” and they all become a little afraid that you’re just going to come after them for lack of a better phrase. It definitely has a negative impact on the climate and the culture of the school, especially with the person that is either working with, it can be toxic. They’re spreading lies or they’re making it out like they’re the victim. It’s not everyone (who) understands what’s happening in the individual teacher’s classrooms. I’ve seen it where it has impacted the climate of the school and if working with that individual is not handled appropriately, they take it personally. It’s definitely a breakdown in the relationship at that point.

However, once the improvement plans concluded, the building climate seemed to improve as staff members ultimately acknowledged the need for remediation and subsequently worked to improve their own practice. Dustin felt the culture improved when effective educators observed underperforming teachers “who weren’t doing what they were supposed to receive the appropriate consequences.” Dustin observed an overall increase in staff professionalism. Jordyn believed the process made other teachers “alert and aware” because they knew the principals would “make sure that you are doing things that are best for kids.” Glen believes the improvement process was “attention getting” and “set a bar for expectations that had not previously been set” by previous school leaders in the school. He reported that high-performing educators expressed their appreciation to him for taking action to address underperformance. Jordyn felt staff respect for her increased because they saw her addressing problems in the school. Jeffrey agreed that staff respect likely increased because the problems had been ongoing and largely were unaddressed:
That was actually a positive for pretty much everybody in the building, because I think everybody in the building saw this and was kind of frustrated that it was allowed to continue for so many years. In some ways, I think that probably increased the respect I had.

Michelle also described a case when the building climate improved during a PDP:

Last year, we had probably our worst performing teacher on a Professional Development Plan and people were thrilled about it because she’s been taking kids out and being a terrible teacher for two decades now. And so finally she had rated low enough that she was on this plan... I mean, people were happy that she was on a plan.

**Outcomes of policy reforms.** Bridges and Groves (1999) defined outcomes as “the particular decisions that have been made with respect to the ground rules, the procedures under which evaluations should be conducted, and the actual evaluations of employee performance.” Participants were asked overarching questions to illuminate their overall opinion of the effectiveness of teacher evaluation policy reform and the value improvement plans. The following subsections describe participant responses when they were asked to reflect on the outcomes of improvement plans and policy reform and the influence of these on the identification of underperforming teachers.

**Improvement plans are a valued lever for educator improvement.** When asked—in hindsight—if they would implement improvement plans or initiate dismissals again, all 20 participants agreed they would. Corey felt the remediation plans were successful in improving classroom practices in his building, but not because the teacher improved their performance. He explained, “Has it resulted and improved instruction? Yes, because those were some of my lowest performing teachers and they are gone!” Harold reflected on when he would use a remediation plan in his school:

I would see myself using the remediation plans for any teacher that is demonstrating that they’re unable to provide sound instruction in the classroom almost immediately. Especially if you’re talking about a (non-tenured) teacher, not allowing them to get to the point where they’re in their third, fourth year. Growing those teachers and providing their
remediation plan. I don’t necessarily believe in automatically letting teachers go. I think if you see something, see the possibilities but they’re struggling in some areas, you provide a remediation plan for them.

Despite the time commitment and workload, Jordyn felt remediation plans were best practice for students:

I think it takes a lot out of the principal. It’s a lot to do when you have so many aspects of the building to run. But I think it’s worth it, because, in this first case, when the moment that it was determined that she was (going) on a plan, I was actually meeting with her; a data meeting. She was not prepared, did not have anything in terms of expectations of what needed to happen. I just told her, “I cannot allow you to be a disservice to kids,” because there were some other things that happened, too. She totally understood. I was like, “I just have to let you know. I’m going to put you on a plan.”

For me, it’s all about kids. I can’t knowingly know that you are not doing what’s best for kids and say that’s okay. No matter how much work it is for me, it is my duty to do that. So yeah, I’d do it again, ’cause I’m giving somebody else who needs improvement, who will probably be on Professional Development Plan after this evaluation cycle. Yeah, I’m doing it again already.

Ann concurred and expanded on the time commitment and workload:

I basically didn’t have a life from March. I truly was not doing my job. My superintendent took over my building. Then I spent that whole summer. I didn’t have a summer that year. It was basically about six months of just constant work, mostly writing and gathering evidence, but I’d do it again. . . .The remediation plan ended up being about 80 typed pages. It was very, very lengthy. No human being could have ever met the requirements of it. We had very, very good legal counsel that guided me. I pretty much stopped teaching or stopped running my building and I was working with the attorneys from the March 1st almost on a daily basis. Just writing and rewriting and presenting things to him and then he’d show it to his counsel.

Ann added that in the future she would “probably ask for time in order to meet all the demands of the building and the special education and discipline and so on.”

Aidan concurred that improvement plans were time consuming, “but if you feel like the end result makes the organization a better place, then you are willing to do it.” He further expanded on the time commitment required:

There is a limit to what an administrator can actually do, because it requires a huge amount of work and a lot of meetings. In my case—attorneys on the phone, assistant
superintendent in my office. It brings it to a whole other level of scrutiny once you do that. But in my case, these were some veteran teachers who were definitely underperforming and I had to make a decision: is the effort I’m going to put in going to result in improvement or removal? And am I willing to take that step?

Despite the workload and time commitment, however, Diana explained how failing to implement an improvement plan is unfair to others:

> It’s not fair to have teachers underperforming in schools. It’s not fair to the children. It’s not fair to the teacher to not know how they’re performing. It certainly is not fair to the other teachers who have to work with them. So yes, I will do remediation and Professional Development Plans if it needs to happen.

Molly shared the personal toll the plan took on her emotionally. The underperforming teacher was, at one time, a close colleague and personal friend when they taught together in the school district. Molly explained,

> In my opinion, it had negative effects on me. I wanted her to succeed and be successful, but I also wanted her to change the way she was teaching, and be a better teacher. In that respect, I was pretty much . . . disheartened or disappointed that she couldn’t make those adjustments. Clearly, they were easy adjustments to make, and they were on the original evaluation tool, so she knew what it was going to be asked. I don’t know. I just don’t understand why she wasn’t able to fix that problem.

> I didn’t enjoy any of it, but by the same token, everybody here knows and even the kids in my classroom, learning is a job, too. You do your job, and everything is fine. If you’re not doing your job, then we’ve got a problem, and I’m that way here in this administrative position, much the same as I was in the classroom as a teacher.

**Mixed perceptions of policy reforms and their local implementation.** All 20 participants believed educators had been evaluated fairly overall, and issues involving bias and fairness were identified, joint committees revised their procedures. For example, after the student growth component was implemented, some joint committees reexamined their procedures and made incremental changes for fairness or consistency. Charlie explained his district first initiated classroom observation procedures that gave an “excellent” classroom observation rating to a teacher with just two or more domains rated “excellent.” Later, the joint committee changed the
scoring criteria and required four domains rated “excellent” in order to receive an “excellent” for the classroom observation portion. Diana described how the Rosedale school district revised their Danielson rubric to improve the alignment and job-fit for early childhood teachers.

However, although the evaluations were perceived as fair, two participants shared how evaluators might resort to “gaming the system” to obtain an overall underperforming rating for a teacher who they perceived as ineffective. Jeffrey described how evaluators must utilize the procedures of evaluation to their advantage:

If you have someone whose professional practices are very questionable, I think you definitely have to be very vigilant in identifying as many different areas (as) “needs improvement” or even “unsatisfactory” just to ensure that overall it comes out . . . as you need to move forward with some steps or plans to help remediate their performance issues.

When assessing the individual components for a teacher who warrants an underperforming rating, Emma explained, “sometimes there’s a numbers game that you feel like you get trapped up in.”

To address other problems of ratings fairness or consistency, participants offered suggestions for changes to their districts’ evaluation processes. For example, Corey recommended the classroom observation domains (two and three in the Danielson Framework) could be weighted more heavily in the summative rating. In his district, the four domains are weighted equally for the professional practice portion of the summative rating. However, Jeffrey offered a contrasting view, as he described how focusing on all four components of the Danielson Framework—intentionally focusing on planning and collaboration—improved collaboration in the Highlane school district.
Participants shared mixed perceptions of the effectiveness of the teacher policy reforms.

Glen explained that local implementation has not met the legislators’ intended goals of the policy:

I think (teachers) understand the spirit of the law, but I think when it comes to practice and putting that law in place, I think it’s met with a lot of resistance and a lot of confusion, and I don’t know if the legislation itself is really accomplishing what we had hoped to accomplish, when including things like student growth.

However, Jordyn and Michelle felt the reforms had a positive effect overall in their districts. Prior to PERA, both participants worked in Austin school district, using an evaluation tool described as a mere “checklist” of teaching behaviors. Jordyn described the old tool as “a couple of boxes to make some comments, but nothing that really gave you an extensive look at that professional practice.” It was difficult to issue a low rating, and the tool did not reflect high-quality teaching performance. After PERA, the Austin school district adopted much of the Danielson framework. Jordyn and Michelle agreed the evaluation changes were positive.

Corey felt policy reforms were not responsible for improved teaching and learning. He explained how another improvement lever—instructional coaching—is more important to teacher improvement than teacher evaluation:

Teaching and learning in my opinion improves as a result of trusted collaboration, whether that’s with your principal or another teacher . . . and experience. So, I’ve been here eight years, we have instructional coaches, and so to me those are the trusted collaborators and we really work hard to make sure that those people have a high level of our time. They are not required to report to me, we keep kind of some walls of separation and I have to trust that what they are doing all day long is working in classrooms with teachers, and providing feedback.

Frank went further, questioning whether evaluation reforms motivated teachers to improve:

I don’t think that most people who are performing at a proficient level or above find any motivation in their rating from the student growth perspective. I don’t think that on a day to day basis people look and say, “You know, if that 30 percent or 40 percent is part of my biannual, yearly summative evaluation is the driving force in me wanting to improve student learning.” I think that it’s something that most people don’t see as a huge
influence. I think if people have been looking to continuously improve in their instruction and help their students learn and grow, then that has always been there. If people have struggled in those areas, having that external force of law isn’t necessarily going to be the thing that changes it.

Jeffrey was unsure about the policy impact on learning and wondered if other school districts had experienced improvement as a result of the reforms.

Stephen explained the levers of reform were well-intentioned, but questioned how they were implemented. He reflected, “Our districts, our communities, our evaluators can use the mechanism to wipe clean all the individuals in education that are underperforming. I don’t necessarily think that’s going to happen.” Vickie discussed the correlation of student growth measures and other achievement tests as failure of the reform:

In my building, based on our test scores and based on where we’ve been for the last six years, my data does not show our school system is getting any better since PERA. Our data has gone down. We’re one of those underperforming poverty districts where looking at student growth is not increasing our student achievement.

Incremental improvements to teaching and learning were attributed to policy reforms. For example, Diana reported that using a teaching framework led to increased higher-order questioning, student ownership, and vocabulary level in her school:

I think it’s turned more learning ownership to the children. Teachers have learned to move to the highest end of the rubric that they have to change their conversations with children and they have to make sure that that vocabulary and the expectations for students is higher.

She elaborated, “PERA could use some work . . . but let’s just get a place to start and that was a good start.” Dustin also credited the teaching framework for describing high-quality teaching his teachers strived to meet:

I would say teaching and learning has improved because teachers are using that framework, the Danielson framework, to get them to kind of make that move from “proficient” to “excellent.” That’s the biggest thing that I’ve seen, and in that (“excellent”) category, so much of that is based on what are the students doing in the class. Our classes are much more student-centered than they were before SB7.
Additional positive outcomes were noted. Ann and Aidan cited improved teacher analysis of classroom data because of the student growth component. Glen felt the reforms “started to change some of the conversations that we have amongst grade level teams, or between myself as an administrator and some teachers, which has been beneficial.” Jeffrey cited collaboration as an improved practice resulting from the reforms. Charlie explained that teaching and learning improved because his school focused on constructivist learning, while using teacher evaluation as a lever to support those efforts:

I would attribute [improvement] . . . to getting away from the whole . . . behaviorist format of instruction into more constructivist and being more student-led. So, I think that focus has pushed teachers a lot more than the evaluation. Then we’ve used the evaluation to solidify and reward where it needs to be rewarded, but also comment where it needs to be commented.

The success of the reforms may be influenced by the attitude and performance level of the teacher prior to implementation. For example, Melody felt strong teachers have improved and were “very happy with the change.” However, her struggling teachers were not “fully on board” and viewed the reforms negatively. Stephen also found high-performing teachers benefitted from the reforms:

Where I really see the change is those teachers that really want to keep that excellent on the Danielson model. They really transform their practices in a more student-centered approaches. They’re setting up their routine. They’re setting up their structures. They’re setting up their classroom environment to really create structures and systems that are student driven, that have student ownership. . . . I see it from those teachers that are self-driven and want that excellence. I think that some of the teachers that are perfectly fine with proficient, that’s not driving them over. It’s not the driving force to change their instructional practice.

Michelle concurred, stating “teachers who want to be continually growing and continually learning appreciate the systems,” while teachers who were “happy with who they are as teachers and want to continue doing exactly what they’ve done year in and year out” preferred the previous system.
Despite the challenges presented by policy mandates, school leaders maintained a positive outlook on the potential of the state’s evaluation reform to improve teaching. Harold explained,

I would say it’s definitely a step in the right direction. Is it the silver bullet? No, but I think it’s part of the process. I think there is still a way to go in terms of providing professional development to teachers about PERA, SB7 and what effective teaching is—what learning looks like for students. I think it’s definitely a good move. I’ve seen improvements around conversations with teachers about what’s happening in classrooms and what is important in the classroom. We still have a way to go, but I think it is definitely a step in the right direction.

Amanda was supportive of evaluation reforms. Reflecting on her leadership practice, she described how she continues to model perseverance through a shared responsibility for learning between the teacher and school leader:

I want everyone to be the best they can be and I want my students to be the best they can be. And I think that like I said, if I can project to my teachers, we are in this together, I am not your superior, I am your leader, I am your cheerleader and our school is our school. Our students are our students. If kindergarten isn’t succeeding, it is not just that kindergarten’s teacher problem, that’s my problem, too. Those are our kindergartners. We need our kindergartners to be the best they can be so that when they go to 1st grade, they can take that next step.

Those students are actually going to live in our communities. We’re going to get their children as our next cycle. So we need everybody to be successful so that our school is success. I am passionate about learning, I’m passionate about success. I grew up in a day and age where teachers had no problem telling you that you’re going to be a nobody. I don’t believe in that, so no teacher is going to say their student can’t learn and I don’t give up on my teacher either. I don’t want anybody to give up on me, so I don’t give up on my teachers.

Participants offered suggestions on professional development supports evaluators need to improve their levels of understanding and implementation of the policy. Emma advocated for more professional development on the meaning of the framework language and coaching staff for improvement. Melody explained the need for professional development in interrater reliability:
I think that the continued, ongoing professional development . . . is huge to be able to collaborate with your evaluators, your co-evaluators, other administrators, just to keep the consistency across the board and to ensure that everyone’s fully understanding all of where they go (on the rubric). . . . We all have natural biases. You don’t want things to come through a bias.

Finally, Glen concurred with Melody, suggesting evaluators need practice and feedback on how they apply the framework to classroom observations:

I think a clear understanding of the tool that we use, whether that be Danielson or something else. I think providing professional development on interrater reliability when it comes to observation as a useful activity, or a professional development opportunity. When we’ve had videos brought in, what we are to rate the person and why and provide evidence, I think it’s always interesting to hear what other evaluators see in a given observation, because inevitably, it’s always a little different from each other. So I think, the more that practice can be done, the more precise our feedback can be to teachers.

Integration of Results

In this section, the results of the quantitative and qualitative phases will be integrated. The quantitative data are examined in alongside the qualitative data to address the influence of micropolitics on the data found in the quantitative phase.

Participant estimates and reported frequencies of underperformance. In the quantitative phase, ISBE data (2017d) were obtained to examine the summative evaluation ratings report to ISBE by respondent districts for the 2016-2017 school year. ISBE reported that, statewide, 97% of educators received an “excellent” or “satisfactory” rating in 2016-2017, leaving 3% of educators identified as underperforming (Table 12). These data were also obtained for the school districts of the qualitative phase participants (ISBE, 2017d). During the qualitative phase, the 20 participants estimated the percentages of underperforming educators in their districts. The data are listed alongside the actual rates of underperformance ratings reported to ISBE in Table 17.
Table 17

Integration of Interview Participants’ Estimates of District Underperformance and Reported District Evaluation Data from 2016-2017

<table>
<thead>
<tr>
<th>Participant</th>
<th>Estimates of underperforming educators in participant districts (%)</th>
<th>Reported underperforming educators in participant districts (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aidan O’Brien</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Amanda Ashbee</td>
<td>&lt; 5%</td>
<td>3%</td>
</tr>
<tr>
<td>Ann Keaton</td>
<td>5-10%</td>
<td>0%</td>
</tr>
<tr>
<td>Charlie Ramirez</td>
<td>10-20%</td>
<td>1%</td>
</tr>
<tr>
<td>Corey Hammer</td>
<td>20%</td>
<td>1%</td>
</tr>
<tr>
<td>Diana Legend</td>
<td>8-10%</td>
<td>1%</td>
</tr>
<tr>
<td>Dustin Ross</td>
<td>15%</td>
<td>0%</td>
</tr>
<tr>
<td>Emma Moore</td>
<td>15%</td>
<td>1%</td>
</tr>
<tr>
<td>Eric Graves (*)</td>
<td>&lt; 5%</td>
<td>0%</td>
</tr>
<tr>
<td>Frank Steele</td>
<td>1-2%</td>
<td>2%</td>
</tr>
<tr>
<td>Glen Tucker</td>
<td>10%</td>
<td>2%</td>
</tr>
<tr>
<td>Harold Cooper</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>Jeffrey Shannon</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Jordyn Harris</td>
<td>No opinion</td>
<td>10%</td>
</tr>
<tr>
<td>Melody Roberts</td>
<td>20%</td>
<td>1%</td>
</tr>
<tr>
<td>Michelle Hale</td>
<td>15%</td>
<td>11%</td>
</tr>
<tr>
<td>Molly Martin</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Rebecca Silver</td>
<td>15-20%</td>
<td>5%</td>
</tr>
<tr>
<td>Stephen Pierce</td>
<td>5-10%</td>
<td>17%</td>
</tr>
<tr>
<td>Vickie Wells</td>
<td>2%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Note. Reported data from ISBE (2017).

The data showed 11 participants underestimated their district’s underperformance rates by 5 percentage points or more than the actual rates of summative evaluation ratings reported to ISBE, with four of those 11 participants underestimating their district’s underperformance rates by 15 percentage points or more. Results from the qualitative phase could explain this discrepancy. First, many participants discussed the problem of summative ratings inflation—noting that too many underperforming educators in their districts received a “satisfactory” or “excellent” rating. Though Jeffrey was within one percentage point of his estimate, his district’s experience with summative ratings inflation may explain why other participants’ estimation did not match the reported summative ratings:

We didn’t have an excessive number of excellent (summative ratings prior to PERA), but then when we did the (student growth) process as a district, they shared with us that for
the first year, they said, “We have 80% of our teachers at excellent,” and that’s not a true reflection of what teachers are like.

Second, summative ratings in other buildings may reflect low interrater reliability throughout the district. For example, in Corey’s school district, the central office administration shared the summative evaluation ratings for each building with all the district principals. Corey recalled,

Seeing an elementary school where all but one or two teachers were rated excellent. When the other elementary schools . . . you see a normal distribution. Two years ago—it was the last time we met about it—in the entire district, one tenured teacher got a “needs improvement.” There is no way that could happen. . . . I think that school principal, for various reasons, was inflating a lot.

Information gained through interviews disclosed many factors that could explain why participants’ estimations of underperformance are not aligned with educator evaluation data submitted to ISBE. Of most significance, the local implementation of PERA and SB7 restricted many participants from issuing underperforming summative ratings. First, the assessments often required minimal effort to earn a “proficient” or “excellent” rating, resulting in inflated student achievement gains. Second, targeted goals were easy to achieve. Third, the results from student growth measures often caused higher summative ratings than the classroom observation ratings. In addition, evaluators issued summative ratings that did not match the classroom performance due to deferral, low interrater reliability in applying the framework language, and potential concerns regarding building climate.

**Professional Development Plans are increasing in use—but are underutilized.** The quantitative data illuminated two significant findings regarding PDP implementation. First, respondents implemented their plans at staggered times between 2010 and 2017. Among respondents, 14.61% had yet to implement the PDP requirement of evaluation reforms. Second, the linear trend showed an increase in the number of educators who started a PDP since 2010. By
2016-2017, one PDP was implemented for every 338 educators working in the respondent school districts.

Among those interviewed, 11 had developed and monitored PDPs during their careers as evaluators in Illinois. These participants described mixed results to implementing PDPs, ranging from successful educator improvement to ongoing underperformance that resulted in separation via Reduction in Force. However, when asked if they would—in hindsight—implement a PDP again, all 11 participants who implemented PDPs responded in the affirmative.

Participants were invited by email to participate in follow-up discussion. Twelve individuals responded and provided their insights on the frequency of PDPs from the 2016-2017 school years as reported in the quantitative phase. Some participants felt the data underestimated the frequency of underperformance due to low identification practices. Aidan stated, “I suppose I’m not surprised at the numbers, yet I also feel they drastically underestimate the number of educators who are in need of significant professional development in order to demonstrate proficiency on the Danielson rubric.” Emma agreed, as she explained the changing needs of educators to improve:

My gut tells me that more than 1 out of every 338 teachers have instructional practices or a lack thereof that would necessitate being put on a PDP plan. . . . I believe during the time in one’s career, there are periods of ebb and flows where we all have practices that dip into basic practice areas where a PDP plan would be beneficial to lay out specific steps to work towards improvement.

Glen described the low frequency as “shameful,” while Michelle and Rebecca agreed more educators should be identified as “needs improvement” and placed on a PDP.

Participants described several reasons for the low reporting numbers. First, the data may have been underreported. Corey speculated the data may not have captured all underperforming educators because some teachers choose retirement or resignation before starting a PDP. Second,
Jeffery questioned if the data underreported educators engaged in PDPs—based upon his reflection of the practices in the Highlane school district:

My experience would definitely suggest otherwise. In my district, in talking with my colleagues I know of several principals that have had at least one teacher on a PDP plan. My first reaction would be wondering how accurate this is given at least my discussions with my colleagues at the elementary level in Highlane.

Second, procedures and other micropolitical factors may deter an evaluator from starting a PDP for an underperforming educator. Melody explained why evaluators might be influenced by workload and procedural challenges:

At first, I was a little shocked that it was what I considered low, but then I thought about it and reflected on mine and colleagues regarding evaluations and Professional Development Plans, it made sense. I truly believe (there) are a lot of administrators (who) will not touch evaluating a certified staff member low to avoid having to process through their district’s policies for PDP.

Next, Aidan reflected on the procedural, workload, and evaluator capacity challenges explain why PDPs are not implemented with more frequency:

Why I’m not surprised is due to a number of factors that may play a role in the process: administrators who fear staff backlash or a negative shift in the building culture if more teachers were on PD plans, administrators who feel overwhelmed by the number of staff members whom they supervise and don’t devote the necessary time to adequately supervise and evaluate teachers, or possibly administrators who don’t feel adequately prepared or have the confidence in their own pedagogy of best practice in instruction.

Finally, Stephen speculated that better hiring and new educator induction practices may have resulted in better educators getting hired. Therefore, the number of PDPs should be low, based on an increase in the hiring of quality educators.

**Remediation plans are rare, but the data reported may paint an inaccurate portrait of improvement efforts.** The quantitative data reported low rates of remediation plan implementation. In the 2016-2017 school year, respondents reported issuing remediation plans at the rate of one for every 1,159 educators. Twelve years earlier, Reeder (2005b) estimated the
ratio was 1 remediation plan for every 930 educators in Illinois. Twelve participants in the qualitative phase had implemented a remediation plan, and when asked if they would—in hindsight—implement a remediation plan again, they unanimously responded in the affirmative.

Twelve principals participating in follow-up discussions reflected on the frequency of remediations from 2016-2017 reported in the quantitative phase. Glen described the remediation data as “shameful.” Emma felt the remediation data did not reflect the range of educator performance she would expect to see in a school:

Within every organization there are teachers at different performance levels. I feel if we are being honest with the current state of education in Illinois in regards to teacher performance and practice, there are more teachers who should be identified as needing a remediation plan to improve their practices and get them to a proficient level. I honestly believe administrators are sometimes afraid to communicate the reality of a teacher’s practices and put the work into the process that a remediation plan entails.

Aidan explained how the data underreports the actual number of educators in need of remediation. He cited lack of central office support, the time and effort required to implement a remediation plan, the fears of educators’ reactions and potential negative effects on the building culture, and deferring the remediation due to retirement as reasons for the low frequency. Jeffery further clarified why remediation plans are rare:

I have a colleague who had a teacher on a remediation plan and I also had a teacher on a remediation plan (who) was eventually released. I understand why this is not popular because it is very intense on the part of principals to collect evidence from multiple observations and requires a secondary evaluator. The process is grueling for evaluators in pulling all of the data together, presumably reviewing it with your HR department to make sure you are accurate—and then all of the meetings with the teacher and union representation.

Corey believed the data underreports identification efforts due to retirements and resignations that may occur prior to the implementation of either plan. Thus, both PDPs and remediation plans may be used as levers by evaluators to encourage underperforming educators to resign or retire to avoid a “needs improvement” or “unsatisfactory” summative evaluation
rating and an improvement plan. Therefore, the data may not credit evaluators accurately for their efforts to employ the improvement strategies included in PERA and SB7 as they work to improve teaching quality.

**Summary**

This chapter shared results of a study on the implementation of evaluation policy reform in Illinois and the influence of micropolitics on the identification of underperforming educators. A questionnaire completed by 89 superintendents found low rates of educators identified as underperforming. However, the data on PDPs showed an increasing trend line since implementation of the reform, partly due to the staggered implementation in districts. Interviews of 20 school principals described how the implementation of student growth measures hindered their ability to issue underperforming summative ratings. The joint committees created procedures that influenced higher summative evaluation ratings that did not reflect the educators’ lower ratings for classroom performance. When identifying an underperforming educator, participants described union influence as neutral to positive, reflecting a shared goal of having quality educators in the classroom. Evaluators deferred low ratings due to factors including staff retirement, avoidance, discomfort in holding difficult conversations, and the workload involved in implementing improvement plans. Despite the time and effort required to implement improvement plans, participants answered affirmatively when asked if they would implement an improvement plan again. Chapter 5 will discuss the conclusions from and implications of the findings reported in this chapter.
Chapter 5

Discussion, Implications, and Recommendations

This chapter begins with a summary of this research study, including a review of the research methodology and major findings. The discussion section expands on the results of the research findings and provides insights for practitioners and scholars regarding potential implications. Additionally, this final chapter concludes with recommendations for practice, policy, and future research in the area of micropolitics of personnel evaluation and education policy implementation.

Overview of Research Methodology

The purpose of this study was to examine the phenomenon of teacher evaluation, focusing on how micropolitics have influenced the implementation of teacher evaluation reforms in Illinois. Two conceptual frameworks framed this study: education policy implementation theory and micropolitics of personnel evaluation. The implementation of education policy is affected by local context, including the policies, people, and places (Honig, 2006). Micropolitics of personnel evaluation provides one lens for examining the interactions of local actors who may influence evaluators when identifying underperforming educators (Bridges & Groves, 1999).

The methodology for this study involved an explanatory sequential mixed methods design with the follow-up explanation variant. In the quantitative phase, personnel and school demographic data from ISBE and 89 public school districts were collected and analyzed. In the qualitative phase, data from semi-structured interviews of 20 principals illuminated the quantitative findings. The following research questions guided this study:
Research question 1. To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools?

Research question 2. How have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms?

Findings

Research question 1. To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools? At the start of this research project, I speculated there would be little or no difference in the identification of underperforming educators before and after the implementation of evaluation reform in Illinois. Although the number of underperforming educators identified remains small overall, trend lines showed small increases in the use of improvement levers and dismissals during implementation of the policy. By 2016-2017, respondents reported 0.296% of educators started a PDP and 0.086% had started a remediation plan. The study found districts implemented their PDPs at staggered times since the passage of PERA, with 14.6% of districts reporting they had not implemented PDPs by 2016-2017. During the 2016-2017 school year, respondents in the reported 0.431% of educators were dismissed. This resulted in a ratio of 1 educator dismissed for every 232 educators in respondent districts.

Research question 2. How have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms? Throughout the qualitative phase, respondents described many factors that influenced the identification of underperforming educators. This information is organized into
four themes for analysis: (a) decisions about procedures of the evaluation plan; (b) decisions by evaluators; (c) actors and their access, interests and power; and (d) outcomes of policy reforms.

Five procedures created by joint committees influenced the identification of underperforming educators. First, 19 of 20 districts adopted the Danielson framework for their evaluation plan. Participants found an evaluation framework helpful in specifying written criteria of teaching performance that could be communicated and understood—with training—by evaluators and educators. Second, Type III assessments showed growth but were easy to score well on. Third, procedures for choosing local assessments led to low-rigor assessments; educators could easily earn the highest student growth scores on these assessments. Fourth, when student growth scores were combined with classroom observation ratings, the student growth score often raised the overall summative rating to a rating higher than the classroom observation rating. As a result, identifying underperforming educators became more difficult, as a high student growth score could move a “needs improvement” classroom observation to a lenient “satisfactory” summative rating overall. Finally, participants experienced additional workload through frequent observation visits, lengthy paperwork requirements, and multiple meetings with educators during the evaluation process.

Five factors influenced evaluators’ decisions when conducting evaluations and implementing the procedures of their district evaluation plan. First, low interrater reliability exists in districts and creates an uneven distribution of summative ratings when evaluators apply different interpretations of the framework and the procedures. Second, evaluators may defer low summative ratings due to an educator’s proximity to retirement and/or their seniority. Third, evaluators may be unwilling to issue a low rating. For example, a marginal educator may earn a higher rating to than deserved to preserve diversity among the school staff. Fourth, evaluators
may use a different improvement lever outside the evaluation plan when deemed more effective or less likely to affect the school culture. Finally, participants discussed the importance of having difficult conversations with educators. Some felt they were easier after evaluation reform, using the framework language to guide and “depersonalize” the conversations.

Several actors were discussed by participants. Parents, students, and school board members provided information to evaluators but did not hinder the identification of underperforming students. In most cases, these actors provided information to the evaluator about underperforming educators, but stopped short of telling the evaluator how to conduct the evaluations. Central office personnel (including superintendents) were supportive, except for one case in which the superintendent blocked the identification, citing danger of legal action against the district. Participants felt the union influence was neutral to positive overall.

Participants described the outcomes of the policy reforms. All participants stated they would initiate an improvement plan again if they needed to do so. Despite the challenge of these plans, participants felt they were best for the school and necessary to implement. However, participants shared mixed feelings about evaluation reforms overall. Some felt the reforms helped learning by defining quality teaching and improving conversations about teaching and learning. Some argued student growth improved educator skills in assessment, but others felt student growth led educators to “game the system” to earn the highest scores. Participants also discussed the need for ongoing professional development for evaluators.

**Integrated findings.** Participants explained the quantitative data might not reflect actual underperformance because evaluators use other means to address underperformance, including negotiated separation agreements and other improvement strategies. In addition, the data may reflect evaluators’ unwillingness to initiate improvement plans due to time, workload, and
potential negative impact on the building climate. When participants were asked to estimate the number of underperforming educators, their estimations often overestimated the frequency of underperformance compared to the actual number of educators identified as underperforming.

**Discussion**

This section contains a discussion of selected findings from this study. In addition, the findings are explored through the lenses of education policy implementation theory (Honig, 2006) and micropolitics of personnel evaluation (Bridges & Groves, 1999).

**Low rates of identification of underperforming educators.** The quantitative data found low rates of underperformance identification in respondent districts. In the PDP data from 2016-2017, when 84.27% of respondent districts had implemented the PDPs, one PDP was implemented for every 338 educators working in the respondent school districts. In the remediation data from 2016-2017, respondent districts implemented one remediation plan for 1,159 educators. In comparison, Reeder (2005b) estimated one educator for every 930 educators received an unsatisfactory rating—resulting in a remediation plan. This finding is important because a decline in remediation plans may be explained by the availability of PDPs. With both PDPs and remediations available after evaluation reform, evaluators may choose to use a PDP instead of a remediation plan. First, a “needs improvement” rating is easier to issue than an “unsatisfactory” rating and uses softer language (Yariv, 2006) than remediation. Second, creating and monitoring a remediation plan is a more lengthy and time consuming process (Dandoy, 2012; Menuey, 2007). Therefore, the PDP is an intermediate method evaluators might employ before implementing a more involved remediation plan.

Respondents in this study dismissed fewer educators annually than has been reported in other states. For example, Thompson (2006) found 0.70% of educators were dismissed in
California. However, analysis of Illinois dismissal data from this study is somewhat problematic. First, non-tenured educators can be released by a non-renewal of their probationary contract (ISBE, 2015). Second, both tenured and non-tenured educators might resign before they are dismissed due to a negotiated settlement or desire to avoid the stigma of dismissal (Reeder, 2005b). Third, the small number of dismissals reported by respondents is highly sensitive to small changes in frequency and may not be generalizable to the entire state. Fourth, the literature on Illinois’ dismissals is limited to those cases that reach a state hearing decision, which is a small subset of educator dismissals statewide (Henry, 2010; London, 1998; Seltzer, 1992; Thompson, 2010). Therefore, although the trend lines showed an increase in dismissals, dismissal data provides an incomplete narrative on the identification of underperforming educators and should be interpreted with caution (London, 1998). These findings highlight the need for more comprehensive and accurate statewide data collection on personnel evaluations so policymakers can study the results of policy implementation in Illinois.

Data from ISBE (2017d) found 3% of all Illinois educators received an underperforming rating in 2016-2017. However, among respondents in the quantitative study, only 1.25% of educators received an underperforming rating in 2016-2017. The statewide results are consistent with the findings from two recent studies. Kraft and Gilmour (2017) combined data from 19 states and reported 2.7% of educators were rated as underperforming. Second, aggregate evaluation data from Michigan from 2011-2012 and 2014-2015 found 97% of Michigan educators earned the two highest summative ratings on a four-rating system (Lenhoff et al., 2017), which is identical to the statewide findings in Illinois from 2016-2017 (ISBE, 2017d). These findings are important because they show the summative ratings distribution in Illinois is similar to that found in other states, including comparable reform efforts in Michigan. Thus,
Despite the efforts to implement evaluation reforms in Illinois, the evaluation ratings are about the same as other states.

**Teachers’ unions generally were supportive of the improvement process.** When unions represented underperforming educators, participants described union leaders as generally neutral to positive in supporting the improvement process. This finding was unanticipated, as I assumed unions were likely to aggressively advocate for the protection of their individual members and fight against the assignment of improvement plans. Unions have been characterized as being distrustful of school leadership, stemming from their history of advocating for basic rights such as job protections against arbitrary dismissal and/or discrimination that are provided in laws for all citizens today (Mead et al., 2012).

When implementing an improvement plan, participants reported the unions and other educators initially may “circle the wagons” to protect their colleague. By the end of an improvement plan, however, participants shared that educators and union leaders were either pleased the educator improved or satisfied that an underperforming colleague was leaving the school. This is an important finding because the literature found evaluators often considered the negative consequences of union interaction before deciding to issue a low summative rating that triggered an improvement plan; evaluators may issue a higher summative rating than may have been justified to avoid union problems (Dandoy, 2012; Mitchell, 2011). In addition, teachers’ unions typically provide rigorous representation and legal defense when educators are facing dismissal or other negative job action, driving the school district to spend significant funds on litigation (Menuey, 2007; Reeder, 2006; Reuland, 2012).

The findings of this study suggest evaluators in Illinois should rethink their concerns about potential objections from union leaders and other educators when addressing
underperformance with an improvement plan. One explanation could be the process used to create the policy reforms. Illinois policy makers intentionally engaged educators and their unions—both locally and in the statewide negotiations—as an intentional part of the change process to increase their support of PERA and SB7 legislation (Regenstein, 2011). Union leaders negotiated and compromised with legislators, advocacy groups, and other stakeholders to avoid harsher outcomes such as the reduction of collective bargaining rights (Lewis, 2011). I speculate the involvement of teachers’ unions in crafting the reform legislation may have created some degree of “buy-in” that has influenced regional and local unions to accept the processes of improvement plans. Since the union helped create the plans, they might feel obligated to carry-out the procedures they agreed to implement. Alternately, Illinois teachers’ unions may have evolved in their organizational goals to promote quality teaching and learning for all students.

Following this study, the United States Supreme Court recently issued a ruling on an important Illinois labor relations case in Janus v. State, County, and Municipal Employees (2018). The court struck down the Illinois law—known as an “agency fee” or “fair share”—that required Illinois governments to deduct collective bargaining fees from the compensation of any employee whose position was covered by the collective bargaining unit but who chose not to join the union. The majority opinion found the First Amendment free speech rights of non-members were violated by withholding these fees without their consent. More research is needed to explore the interaction of union leaders and educator improvement plans in Illinois and other states and whether the Janus decision will influence the power balance between school boards and teachers’ unions.

**Student growth calculations influenced the inflation of summative ratings.** Many participants found local procedures for student growth measures made it challenging to assign an
underperforming educator a low summative rating. First, participants shared it was easy for educators to earn the highest score on their student growth measures, which is consistent with the findings from a study of evaluation reform implementation in the Chicago Public Schools. The study found teacher-created and teacher-graded assessments provided relatively simple, unchallenging pathways for educators to earn the highest scores on their student growth assessments—especially assessments measuring simple student growth (Jiang et al., 2015). In contrast, Darling-Hammond (2013) explained that evaluation plans should not use grade-level standards (such as classroom assessments) to measure student growth. She advocated for continuous scale assessments to show growth above and below grade levels. In addition, participants perceived educators could “game” the system by using the procedures to achieve high student growth scores. This finding was consistent with the Chicago Public Schools study in which educators also perceived the student growth measures as easy to “game” by simply assigning students a low score on the first assessment and giving them a higher score on the second assessment (Jiang et al., 2015). These findings are important because inflated summative ratings may explain the low rates of identification of underperforming educators.

**Conversations about teaching and learning improved following the reforms.** Nine participants described how conversations with teachers improved following the reform. This finding highlights a positive outcome of evaluation reform and is supported by Danielson’s vision of her teaching framework—defining good teaching practices to facilitate professional learning conversations between the educator and the evaluator (2010). Because of reform, all evaluators completed training to become a qualified evaluator, including instruction on professional conversations with teachers in the pre- and post-conference model. In this study, all 20 participants worked in districts that adopted a teaching framework as required by reform, and
19 of those districts used the state-recommended Danielson framework (ISBE, 2015). Frank attributed the improved conversations to the Danielson teaching framework, as it provided “definite observable actions and the framework for people to be able to have conversations to improve.” This finding is important because it highlights a positive outcome of evaluation reform—with the potential to positively influence student learning. Hattie (2009) found formative feedback to teachers had a .90 effect size on student achievement, and Tuytens and Devos (2011) found feedback offered in post-conferences leads to greater teacher engagement in improvement efforts.

*The problematic nature of evaluation consistency.* Participants described challenges with inconsistent application of their evaluation plan. First, evaluators may not apply the rubric language with consistency or fidelity to the written criteria in the district framework. Second, some evaluators may have chosen to rate educators higher to avoid conflicts with individual educators and schoolwide conflicts and perceived discomfort that may result from low summative ratings and the monitoring of improvement plans. This finding of lack of consistency is consistent with the literature. Evaluators within the same school or district may vary in their application of evaluation criteria and/or procedures (Bernstein, 2004). Some may use softer language when providing feedback to avoid conflicts with staff or because of personal discomfort with providing candid, yet constructive feedback (Yariv, 2006). Also, the evaluator may decide to defer low ratings or be unwilling to assign low ratings. These decisions might be influenced by the educator’s seniority or retirement status, the evaluator’s discomfort with difficult conversations about improvement (Mead et al., 2012), or the evaluator’s lack of confidence in the effectiveness of improvement plans (Kraft & Gilmour, 2016).
These findings are significant for two reasons. First, summative evaluation ratings are used to determine Reduction in Force layoff sequences across the district (ISBE, 2015). If the evaluation plan is not implemented consistently by all evaluators, some educators could be ranked too high or too low on the layoff sequence. Second, if evaluators fail to rate underperforming educators with a “needs improvement” or “unsatisfactory” summative rating, the educator will not benefit from an improvement plan to address their performance.

**Teacher evaluation reforms brought about increased workloads for supervision, evaluation, and improvement plans.** Participants described additional responsibilities and time pressures that resulted from policy reforms. Additional conferences with educators and increased frequency of formal and informal observations were described. These findings were consistent with the literature on evaluation reforms, which has noted that evaluation reforms, in general, have required evaluators to devote more time to evaluation processes (Donaldson & Papay, 2014; Drake et al., 2016) and created additional work hours for evaluators (Lavigne & Chamberlain, 2017). In addition, evaluators who chose to implement improvement levers were challenged with additional hours and paperwork to create and implement improvement plans with underperforming educators (Kraft & Gilmour, 2017). This finding reinforces the need to reexamine the role and workload of school leaders (Grubb & Flessa, 2006). Without changes, evaluators are less likely to engage underperforming educators in improvement plans.

**Evaluators’ perceptions of underperformance rates do not match reality.** Participants were asked to estimate the actual percentage of teachers who were underperforming in their school district. The study found 11 participants overestimated the actual percentage by 5 percentage points or more when compared to the actual district evaluation rates reported by ISBE (2017d). The teacher evaluation literature also has reported wide gaps in administrators’
perceptions of underperformance rates and the actual rates of identification. For example, Kraft and Gilmour (2017) found 6.5% of educators were rated as underperforming in one district, but evaluators estimated 27.8% of educators in their district were underperforming. Tucker (1997) suggested underreporting of ineffective educators within districts was caused by inflation of summative evaluation ratings. This finding is important because it suggests far more Illinois educators are underperforming than has been reported by the statewide summative evaluation ratings. These underperforming educators continue to teach in Illinois schools and engage students without the assistance of an improvement plan.

**The local context matters when implementing education policy.** This section discusses the findings through the lens of education policy implementation theory. The implementation of education policy is affected by local context, including the policies, people, and places (Honig, 2006). In the context of this study, these three factors interacted to influence the implementation of evaluation reforms. Because the reforms relied heavily on local implementation, this theory was a useful framework to study how local context shaped the implementation and why implementation looks different in each school district. Three subsections describe the findings through three factors of education policy implementation.

**Unions and their joint committee members sought evaluation procedures favorable to their positions.** The people who are responsible for the applications of policy in their organizations exert a large influence in shaping policy implementation (Honig, 2006). In the context of this study, joint committees created the local procedures to implement evaluation policy reforms at the local school district level. These committees included equal representation from the administration and teachers’ union. However, participants in this study spoke more often about the influence of educators on the joint committee. Participants described how some
educators on joint committees were defensive to protect educators from low summative ratings. This is consistent with the findings of Conley and Glasman (2008) regarding educators’ fears of the evaluation process and potential employment loss. In addition, some participants described educators who protected their interests by shaping procedures favorable to educators. These procedures included favorable procedures for student growth measures and the calculation methods for combining student growth scores and classroom performance rating for the summative evaluation. However, two participants described a collaborative joint committee process. These findings are important because understanding the motivations of people who shaped procedures helps to explain the outcomes. However, the framework could be strengthened by understanding more about the influence of relationships between the joint committee members in the outcome of the committee’s work.

*The goals of evaluation policy were influenced by local joint committees whose educator members focused on their own interests over student learning.* Policy can be described by the dimensions of goals, targets, and tools (Honig, 2006). In the context of this study, the broad goals of evaluation reform included increasing accountability and improving educator quality. The goals of Illinois evaluation reforms were influenced by the targets of reform—educators—who used the policy tools of local implementation and the joint committee.

Participants generally perceived that joint committees were not focused on the goal of improved student learning as they implemented student growth measures. Instead, participants reported the joint committees developed procedures that resulted in minimal accountability for educators to improve student learning. These findings are consistent with research by Odden (1991), who explained state policy reforms have the potential to influence change at the local level, but he predicted most changes were unlikely to produce gains in student achievement. In
addition, Owen (2006) suggested that adult concerns regarding their own personal and professional needs would supersede the interests of student learning and classrooms in school micropolitics. These findings are important because they explain how educators on the joint committee used the tools of reform to shape reflect the interests of educators over the goal of improving student learning.

*Illinois was a challenging location to negotiate and implement evaluation reform.* The site of policy implementation influences the outcomes of policy (Honig, 2006). Because Illinois has a long history of collective bargaining and influential teachers’ unions, evaluation reforms included union participation and voice in the crafting of legislation and the implementation of local procedures (Regenstein, 2011). This study found that local implementation often resulted in teacher-friendly evaluation procedures—with less focus on student learning than advocated by groups such as Stand for Children and Advance Illinois. This finding is aligned with reports from critics who predicted little change would result from teacher evaluation reforms (Russo, 2011). These findings are important to understand the historical labor context and why the reforms were shaped to include teachers’ unions in shaping local implementation.

*Micropolitics influenced the implementation of evaluation reform.* This section applies the lens of micropolitics of personnel evaluation to examine the decisions, influences of local actors, and outcomes of the decisions when identifying underperforming educators (Bridges & Groves, 1999). Three subsections will discuss the findings through these factors.

*The actions of local joint committees were designed to protect the interests of educators.* In the context of this study, the first level of decisions in this framework are the ground rules established by legislation and regulation. The second level of decisions are the procedures left to the local school district and its collective bargaining agent for implementation.
This study found educators used their power on some joint committees to influence procedures to protect their employment interests (Conley & Glasman, 2008). Participants explained how student growth processes in some districts thwarted their ability to identify underperforming teachers. The third level of decisions are the evaluations. Consistent with the literature, participants shared many reasons why they might issue a summative rating that inaccurately reports the educators’ performance. In this context, evaluators hindered the implementation of reform by their own evaluation decisions. Participants also shared success stories about decisions to implement improvement plans and the confidence in making that same decision again.

In the context of this study, the micropolitics of personnel evaluation framework provided structure for the three kinds of decisions three levels of decisions impacting evaluation reform. However, it could be strengthened by including the new model of joint committees in Illinois. While some might classify joint committees as a collective bargaining process, the intent of the law was a nuanced collaboration model with educators and administrators holding equal power (Regenstein, 2011).

Once the evaluation plan procedures were created, actors rarely interfered with evaluators’ duties when identifying underperforming educators. In the context of this study, the actors are the school board members, parents, teachers, students, superintendents and central office administrators, and educators. These actors influenced the identification of underperforming educators, but none were identified as hindering underperformance identification. In contrast, school board members, parents, and students engaged with evaluators to support their work. These actors sometimes provided information that was helpful to the evaluators, but they made no attempt to use their power to influence the evaluation process and
hinder underperformance identification. More research is needed to study and understand the role of school boards, students, and parents in identifying underperforming educators.

Superintendents and central office administrators were helpful in providing resource supports, especially when the district leadership communicated a strong culture and vision of quality teaching. This support is consistent with prior literature on district culture (Bridges & Groves, 1999; Thompson, 2006). Participants stated their central office leaders and superintendent, in general, did not hinder their efforts to identify underperforming educators, except in one instance when an evaluator’s desire to identify two underperforming educators was rejected by the superintendent, citing concerns about potential litigation. Concerns about litigation costs are consistent with previous reports (Reeder, 2006; Reuland, 2012).

Finally, educators influenced the building climate. In contrast to previous literature on the evaluation process (Thompson, 2006), many principals described how their building climates actually improved during and/or after implementation of an improvement plan—especially when underperforming educators separated from the district or improved their performance. Although educators may form coalitions early in the process in response to identification (Bridges & Groves, 1999), participants described how educators began to understand the need for improvement, even sharing strategic information with principals to inform them about problems with an underperforming teacher. More research is needed to understand this phenomenon.

In this study, the micropolitics of personnel evaluation framework was useful for defining the actors involved in the evaluation process and noting their interactions. However, the framework could be strengthened to include how coalitions of educators act to either support or oppose the identified teacher. Given the findings in this study that unions were neutral to
supportive during improvement plans, the framework could explore how individuals and/or coalitions act to support underperformance identification.

The outcomes of evaluation policy reform reflected limited use of improvement levers.

Findings from the quantitative portion of this study showed that in respondent districts, most educators earned the highest evaluation ratings and the use of improvement levers continues to be limited. This finding is consistent with previous research conducted in Michigan (Lenhoff et al., 2017) and in 19 other states (Kraft & Gilmour, 2017). In addition, summative ratings inflation in Illinois was evidenced by the discrepancies of participants’ estimates of underperformance and the actual number of educators identified as underperforming; this finding is consistent with that found in other studies (Kraft & Gilmour, 2017; Thompson, 2006; Tucker 1997). Several factors explain this discrepancy in Illinois. First, some underperforming educators could not be identified for an improvement plan because their student growth scores led to higher summative evaluation scores (Jiang et al., 2015). Second, evaluators may be unwilling to assign educators low ratings due to their seniority, impending retirement, and other factors. Finally, evaluators may avoid ratings that trigger an improvement plan because of the administrative time and workload requirements of an improvement plan (Donaldson & Papay, 2014; Drake et al., 2016). These findings are important because Illinois policy makers need to understand the reasons why implementation of improvement levers was low.

Implications

Teacher evaluation policy continues to be an evolving topic for scholars, practitioners, and policymakers. This study is important because it investigated the implementation of recent evaluation reforms in Illinois to determine whether the reforms met their intended outcomes. In addition, this study contributes to our understanding of how micropolitical factors influenced the
identification of underperforming educators in Illinois public schools following evaluation policy reform. This section will discuss three implications of this study.

First, participants described problems with the local implementation of the required student growth component. Although the legislators’ intent for the PDP lever was to improve teaching performance, this study found that local school districts developed processes that significantly diminished the influence of the student growth factor. When educators received inflated summative evaluation ratings that did not accurately reflect their performance, evaluators were challenged to implement an improvement lever. For example, if an educator received a “needs improvement” or “unsatisfactory” on the professional practice component but their student growth score raised their overall performance rating to a “satisfactory” summative evaluation rating, the educator would not benefit from the improvement opportunities and accountability of a formal PDP. Because of this challenge, an underperforming educator will remain in the classroom with no requirement to improve his teaching and learning practices and thus may cause long-term harm to students and student learning (Rivkin et al., 2005).

Second, when administrators do not evaluate with consistency within their buildings and/or across their district, educators may question the value and credibility of the evaluation process. The effectiveness of the evaluation process may decline and educators may not fully engage with the process to improve their own performance and student learning. When educators observe lenient evaluation standards in one building but more stringent standards in another, their frustrations could prompt a loss in confidence of the fidelity of classroom observations by evaluators. If this confidence is lost, it could influence the trusting relationships that are necessary between educators and evaluators so that productive professional conversations and formative feedback can occur (Copland & Knapp, 2006; Fullan, 2014). Educators desire specific
feedback and consistent ratings (Bernstein, 2004). However, the lost trust could reduce educators’ receptiveness to feedback on their teaching performance (Mead et al., 2012), as well as their motivations to improve their practice.

Third, Illinois schools demonstrated low rates of using formal improvement levers. State legislation created levers to improve teaching performance by change the procedures to make it easier to remove underperforming educators from the classroom when they failed to improve (Regenstein, 2011). However, if administrators fail to employ these levers, underperforming educators potentially will remain in the classrooms and continue to negatively affect student learning (Rivkin et al., 2005). This is important because having a high-quality educator is vital to students’ success (Guarino et al., 2006; Little & Miller, 2007).

**Recommendations for Policy and Practice**

This section contains recommendations for policymakers and educational practitioners. These recommendations may be useful for others interested in understanding how micropolitics influences the identification of underperforming educators. However, the results should be used with professional judgment. First, Illinois is a collective bargaining state with a strong history of influential teachers’ unions. The PERA and SB7 legislation negotiation included representatives from both statewide teachers’ unions whose voices were influential in shaping policy reforms (Regenstein, 2011). Second, local joint committees were given wide latitude to implement local evaluation plans. Participants noted how local plans were influenced by educators on these committees who were protecting their interests in achieving high summative evaluation ratings and continued job employment. Third, Illinois has an extensive history of tenure and job protections for educators that may not exist in other states.
In addition, the findings from this study must be viewed in the context of its limitations. First, member checks were conducted post-analysis. Second, a small number of school districts were represented by the quantitative data. Only 10.36% of Illinois public school districts completed the questionnaire, comprising only 6.37% of educators and 5.90% of students in Illinois public school. The quantitative data did not include the three largest school districts in Illinois. Second, the qualitative phase only interviewed 20 principals from Illinois schools. To be included in the study, participants were required to have at least seven years of experience evaluating in Illinois public schools and experience with at least one improvement lever during their career. Therefore, these findings may not be generalizable to all evaluators, educators, and school districts in the state.

**Recommendations for policy.** PERA and SB7 represented significant shifts in statewide teacher evaluation policy. The implementation of this legislation has now provided an opportunity for policymakers to reflect on the outcomes of the policies, to determine whether the policies are having their intended effects, and to note whether the policies are meeting the needs of public schools in Illinois. Four recommendations for policy are presented in this section.

*Policymakers should consider the value of student growth measures in achieving the intended outcomes of teacher evaluation reform.* The data showed student growth measures created additional responsibilities and added workloads for educators and students. Although the original goal was to create expanded accountability levels for educators, the local implementation of student growth appears to have had a different effect in some districts. Nine participants reported that their districts approved student assessments that required minimal effort to score a “proficient” or “excellent” student growth rating, while 15 participants described procedures that resulted in inflated summative ratings that are unmatched to educator
performance. In addition, in some instances the influence of student growth on summative ratings hindered evaluators from identifying underperforming educators. These unintended consequences have served to undermine the original goal of accountability. Consistent with extant research, Popham (2007) suggested poorly designed educator accountability systems could have a negative effect on the goal to improve educators. Murphy (2013) also questioned the overall effects of evaluation reforms, asking whether the additional time and efforts were worthwhile. Therefore, policy changes should be considered, such as untethering student growth scores from the summative evaluation rating and/or increasing the rigor of the assessments and scoring criteria.

_Evaluators and joint committees need additional support in designing and implementing the student growth component._ First, members of joint committees and evaluators would benefit from training on the development and approval of student growth measures. Participants in this study described problems with procedures for combining classroom performance ratings with student growth scores to calculate the summative rating, as well as challenges with assessments that required minimal effort to show growth and earn a “proficient” or “excellent” rating. Unlike evaluators, joint committee members were not required to participate in training when reforms were enacted (ISBE, 2015). Because joint committees establish the local procedures for educator evaluation, engaging those committee member and the evaluators in conversations and professional learning about assessment practices may influence joint committees to reexamine and improve their local procedures for student growth measures. Evaluators also need support in approving student growth assessments and coaching educators in assessment. Evaluators would benefit from training to improve their knowledge of
assessment design principals to improve their confidence in approving assessments that challenge educators to improve student learning.

**Evaluators need additional training on conducting classroom observations.** Reliability and consistency of evaluation processes and ratings within and across districts could be improved with additional evaluator training (Goldrick et al., 2013; Weisberg et al., 2009). Four participants described problems with interrater reliability in their districts. Participants in this study asked for more opportunities to observe live classroom settings with their peers to assess classroom performance and discuss with their administrative colleagues where their observation scores fall on the evaluation framework, so they could facilitate interrater reliability within their districts, and ideally throughout the state. This training could be offered regionally—throughout the school year—at host sites when students are in attendance. Removing this training from the responsibility of local school districts provides a space to define consistent teaching standards for all schools statewide, as opposed to local schools calibrating their own local vision of teaching quality. This training could complement the mandatory training modules required approximately five years after Illinois evaluators initially are approved to conduct observations.

**Implement annual statewide data collection from school districts to regularly review the distribution of all evaluation ratings and the frequency of teacher improvement levers.** ISBE has initiated the PERA requirement to collect limited evaluation data on educator performance, but only the aggregate quantity of “proficient” and “excellent” summative ratings is reported on a public website (ISBE, 2017d). This study was limited by a low number of superintendents who responded to the questionnaire. Respondents represented just 10.36% of public school districts, 5.90% of students, and 6.37% of educators statewide. Having comprehensive personnel data would help Illinois policymakers understand the effects of
evaluation policy reforms at the local level—knowing whether Illinois school districts are using the statutory improvement levers provided within the school code to identify underperforming educators and seek to either improve their practices or dismiss them from employment. If district leaders were required to report this data annually, state education officials and policymakers could track the frequency of improvement plans, similar to the statutory process for the collection of summative evaluation ratings under PERA.

**Recommendations for practice.** The primary focus of this study was on the implementation of teacher evaluation policy reforms in Illinois. Therefore, the findings of this study serve best to inform the practice of schools and school leaders in Illinois. The findings specifically focused on local implementation of teacher evaluation polices. Five recommendations for practice are presented in this section.

**District leadership must articulate and reinforce clear, rigorous standards for educator performance and adhere to them—regardless of political pressures.** The expectations for educator quality vary from district to district. Despite statewide interrater training on the Danielson framework, evaluators and districts assess educators with differing levels of rigor and apply the evaluation framework inconsistently. Evaluators would benefit from a shared vision of teaching practice (Darling-Hammond, 2013), and it is incumbent upon central office administrators to ensure that these standards are consistently applied in schools throughout their districts (Bridges & Groves, 1999). With 19 participants describing positive support from superintendents and central office personnel, Vickie emphasized the importance for clear district leadership to support the work of evaluators in identifying underperformance. Because her superintendent shared a clear vision of improving teachers’ performance, several improvement plans had been implemented in her district.
Create mechanisms to support evaluators when implementation of a Professional Development Plan or remediation plan is warranted. The job of school leadership is increasingly complex, challenging, and time-consuming (Grubb & Flessa, 2006). However, the time commitment and workload of an improvement plan can be a barrier that affects the leader’s ability to fully carry out all responsibilities inherent within their job descriptions. Eleven participants described an increased workload following the Illinois teacher evaluation reforms, including the challenge of managing a time-consuming improvement plan. Flexible and creative assistance—such as providing job release time, hiring a temporary administrator to assume other duties, distributing leadership responsibilities to both formal and informal leaders, and/or assigning additional staff—should be considered to support and encourage evaluators in the challenging work of educator improvement.

Participate in intentional professional development experiences to maintain and enhance evaluators’ skills. Evaluators benefit from training and support to improve their evaluation skills (Darling-Hammond, 2013; Goldrick et al., 2013; Weisberg et al., 2009). First, evaluators should continue to improve their knowledge and understanding of the Danielson Framework for Teaching (Danielson, 2013) and the research supporting the framework. Second, evaluators should collaboratively experience live classroom observations to increase their interrater reliability within their schools and districts. Meetings after these observations should include peer discussions led by a master evaluator skilled in collecting observational data, aligning observational notes with the teaching framework, and leading learner-centered conversations about teaching and learning. Finally, evaluators should build their capacity to hold difficult and challenging conversations with underperforming educators. The study found professional learning conversations improved since the reforms, and providing continued to
support to evaluators is essential. Central office administrators can provide professional
development opportunities designed to assist evaluators with identifying areas of
underperformance and surfacing these concerns with their faculty

Reexamine the student growth assessments and evaluation plan to refocus on
increasing student learning. Fifteen participants described problems with the procedures for
calculating summative ratings with student growth scores, leading to ratings inflation that
hindered the ability to identify underperformance. Educators and evaluators should focus on the
intended goals of the legislation: improving student learning and improving educator
effectiveness. As the policy has been implemented in local school districts, many educators are
focused on the summative evaluation rating rather than using the process to reflect on their
teaching practices and to improve student learning. Conversations with the joint committee and
among evaluators in districts are crucial to developing a shared vision of how the evaluation
process can be an effect mechanism to improve student learning. Student growth assessments
should be reexamined for their ability to influence student achievement and educator quality,
with revisions made for approval by the school district’s joint committee and/or evaluator.

The procedures for determining the ratings for student growth scores and weighting of
student growth scores in the summative rating should be studied to ensure the classroom
observation ratings will drive the summative ratings for underperforming educators. As
examples, districts could explore evaluation procedures that assign an overall “needs
improvement” on the summative evaluation if one or more domains are rated as “needs
improvement.” Districts could also explore scoring procedures to give more weight to domains
based on their direct relevance to classroom performance.
Ensure evaluators have sufficient time and support to conduct fair and thorough evaluations. Because of the increasing complexity of the principalship, many evaluators are consumed with a myriad of non-instructional duties and must work longer hours to complete the time-intensive evaluation process (Kraft & Gilmour, 2017; Lavigne & Chamberlain, 2017). However, the role of the principal has evolved within the last few decades to intensively focus on student learning. Therefore, alternative leadership models—such as the models discussed by Grubb and Flessa (2006)—should be explored for their potential to increase evaluators’ ability to focus on teaching and learning. These models could include dual or co-principalships and/or distributing leadership responsibilities to staff committees and teacher leaders. By reducing their management tasks and other non-instructional responsibilities, school leaders will have more time to focus on student learning and educator improvement through supervision and evaluation.

All districts must need access to all the improvement levers afforded to them by the policy. The study found nearly 15% of respondent districts had not implemented the required PDP component by 2016-2017. Yet, participants found PDPs were a useful intermediate tool for identifying underperforming educators. Policymakers must hold all school districts accountable for implementing the law, while also supporting districts needing assistance with creating their PDP processes.

Recommendations for Additional Research

Four recommendations for future research are presented in this section.

Conduct follow-up studies of evaluation data with more respondents for analysis of longitudinal data. This study was limited by the small number of districts that provided responses to the questionnaire. If fewer years of data—and more recent data—were requested, perhaps more school districts would choose to participate in a voluntary research study. These
findings could be compared to the data from this study to more fully investigate changes in districts’ rates of identifying underperforming educators.

**Interview central office leadership—including superintendents—to provide additional information to expand upon the building-level perspectives of school administrators.** Central office leaders provide crucial leadership that establishes a vision for policy implementation in their districts and, arguably, they are responsible to ensure that building administrators are trained and socialized into the expectations for evaluating educators and holding them accountable for their professional practices within the districts. Therefore, their perspectives could provide valuable insights. Central office leaders likely were involved in the work of joint committees who established the processes and procedures for evaluation policy implementation in their districts. Their insights could further illuminate our understanding of the micropolitical factors of the early implementation through the joint committees. Second, the perceptions of central office leaders could provide additional perspectives on the work of building-level evaluators and the supports evaluators need from central office leaders to identify underperforming educators.

**Study the influence of improvement plans on the school culture—before, during, and after the implementation of a plan for an underperforming educator.** Several principals in this study described how addressing underperformance can improve the building culture, as other educators become aware that their administration is taking the appropriate steps to hold educators accountable and ensure that students have positive learning experiences. Filling this research gap could provide information to support evaluators considering the micropolitical consequences—both positive and negative—of implementing an improvement plan for an underperforming educator in their school. This research could illuminate strategies to mitigate
potential negative effects of improvement levers in schools. This research could expand upon the findings of this study regarding the responses of unions and educators when educator underperformance is addressed in their buildings. Researchers could study whether engaging the unions early in the policy creation process influenced these unions and their members to “buy in” to the implementation of the policies.

**Expand upon the findings of the qualitative phase through survey research.** The generalizability of this study is limited by the low number of respondents in the quantitative phase \( N = 89 \) and the small number of participants interviewed in the qualitative phase \( N = 20 \). However, statewide or national survey research could examine whether the findings of the qualitative phase could be generalized to a wider population. The questionnaire could investigate the perceptions of evaluators on the strength or influence of the findings of this study across more districts statewide than were reached in this study.

**Conclusion**

Teacher evaluation reform remains a challenging political issue for public schools in Illinois, as well as throughout the nation. These legislative reforms were made with expectations for administrators to improve teaching and learning in public schools. This study was significant because of the need to evaluate the implementation of Illinois education policy reforms to determine if they met their intended outcomes. This mixed methods study investigated the phenomenon of teacher evaluation, focusing on how micropolitics have influenced the implementation of teacher evaluation reforms in Illinois. The study was framed by two conceptual frameworks: micropolitics of personnel evaluation (Bridges & Groves, 1999) and education policy implementation theory (Honig, 2006). This study fills a gap in the literature by providing quantitative data on the number of educator improvement plans issued in Illinois. In
addition, the voices of school principals provide policymakers with information to determine the direction for future policy changes. This study found decisions made by joint committees at the local school district level resulted in scoring procedures that inflated summative evaluation ratings. As a result, some evaluators were challenged to identify underperforming educators with a “needs improvement” or “unsatisfactory” rating—the ratings needed for a PDP or remediation plan. In addition, consistent with data from other states, low rates of educators were identified for improvement plans or dismissals. Illinois policymakers should review these findings to determine whether changes are necessary to meet the intended goals of the evaluation reform legislation.
References


Educational Administration Quarterly, 29, 142-163. doi:10.1177/0013161X93029002003

Educational Administration Quarterly, 38, 6-44. doi:10.1177/0013161X02381002


Cameron, R. (2009). A sequential mixed model research design: Design, analytical and display issues. *International Journal of Multiple Research Approaches, 3*, 140-152. doi:10.5172/mra.3.2.140


systems: Making the most of multiple measures (pp. 116-130). New York, NY: Teachers College Press.


175


Ivankova, N. V. (2014). Implementing quality criteria in designing and conducting a sequential
QUAN → QUAL mixed methods study of student engagement with learning applied
research methods online. *Journal of Mixed Methods Research, 8*(1), 25-51.
doi:10.1177/1558689813487945

explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20.
doi:10.1177/1525822X05282260

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *Future of


Jones, A. C. (1985). *Content analysis of teacher dismissal cases for incompetence under the
Illinois Tenure Teacher Hearing Officer Act, 1975-1983* (Doctoral dissertation, Loyola


Reuland, A. (2012). *The differences between novice and experienced public middle school principals in the decision to remediate a tenured teacher* (Doctoral dissertation, Loyola


University at Carbondale). Available from ProQuest Dissertations and Theses. (UMI No. 9239746)


Retrieved from https://igpa.uillinois.edu/sites/igpa.uillinois.edu/files/reports/IR12-Ch7_edpolicyW.pdf


## Appendix A

### Data Collection Matrix

<table>
<thead>
<tr>
<th>Question</th>
<th>Analysis</th>
<th>Source</th>
<th>Research Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1: To what extent has the implementation of teacher evaluation reforms affected the frequency of identifying underperforming teachers in Illinois public schools?</td>
<td>Tables and graphs, descriptive statistics</td>
<td>FOIA request to ISBE and questionnaires of school districts</td>
<td>Quantitative</td>
</tr>
<tr>
<td>RQ2: How have micropolitical factors influenced principals in the identification of underperforming teachers in Illinois since the implementation of teacher evaluation reforms?</td>
<td>Interviews of principals to identify codes and themes</td>
<td>Interviews of principals</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>
Appendix B

Visual Model of Study

<table>
<thead>
<tr>
<th>Phase</th>
<th>Procedure</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>quant Data Collection</td>
<td>- FOIA requests from ISBE, to obtain school district contact information&lt;br&gt;- Survey data collection from Illinois school districts in Illinois (N = 859)</td>
<td>Numeric data on remediations, PDP, dismissals</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>quant Data Analysis</td>
<td>- Frequency counts&lt;br&gt;- Descriptive statistics&lt;br&gt;- Microsoft Excel 2013 software</td>
<td>Descriptive statistics&lt;br&gt;Tables and graphs&lt;br&gt;Discussion of findings</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>Connecting quant and qual</td>
<td>- Determine if data warrant further study&lt;br&gt;- Purposeful sampling to select participants for cases&lt;br&gt;- Develop interview questions</td>
<td>Interview protocol</td>
</tr>
<tr>
<td>phases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qual Data Collection</td>
<td>- In-depth semi-structured interviews (N = 20)&lt;br&gt;- Follow-up questions</td>
<td>Text data (interview transcripts)</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>qual Data Analysis</td>
<td>- Coding and thematic analysis&lt;br&gt;- HyperRESEARCH software v. 3.7.3&lt;br&gt;- Member checking</td>
<td>Codes and themes&lt;br&gt;Discussion of findings</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>Integration of quant and</td>
<td>- Follow-up questions to participants&lt;br&gt;- Interpretation and explanation of the quantitative and qualitative results</td>
<td>Discussion&lt;br&gt;Implications&lt;br&gt;Future research</td>
</tr>
<tr>
<td>qual Results</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Model adapted from Ivankova et al. (2006).
Appendix C

Advanced Survey Request

1050 Shawnee Court
Bourbonnais, IL 60914
September 25, 2017

Dear Illinois Public School Superintendent

Re: SURVEY ON TEACHER EVALUATION

I am a Doctor of Education degree candidate in Education Policy, Organization and Leadership at the University of Illinois at Urbana-Champaign and am gathering data for a dissertation research study entitled *A Mixed Methods Study of Teacher Evaluation and Micropolitics in Illinois*. The purpose of this study is to explore how micropolitics influences qualified evaluators when identifying underperforming teachers. Dr. Donald Hackmann, Professor, is my dissertation director.

On Wednesday, October 4, I will send you a link to a survey. The survey will request data regarding certificated teaching staff in your school district, including teachers, school psychologists, social workers, school counselors, and school librarians. Please exclude school administrators, such as superintendents, principals, assistant principals, and curriculum directors. Only include deans of students or other quasi-administrative positions when these positions are evaluated under the same conditions as certificated teaching staff. Exclude paraprofessionals, cooks, administrative assistants, and other non-certificated positions.

The survey will require district office personnel to gather information. To help you prepare, I am attaching a PDF file with the survey. When compiling the data, please provide as much data as you are able to. For the purposes of this research, we prefer as much data as possible but will accept whatever data you can provide. The survey will open on Wednesday, October 4 and close on Friday, October 20.

If you have any questions, please contact me at (XXX) XXX-XXXX or dconr2@illinois.edu. Alternatively, you may contact Dr. Donald Hackmann at phone (217) 333-0230 or dghack@illinois.edu.

Thank you for your assistance.

Sincerely,

David Conrad
University of Illinois
Doctoral Student in Education Policy, Organization and Leadership
1050 Shawnee Court
Bourbonnais, IL 60914
dconr2@illinois.edu
### Appendix D

**Quantitative Phase Aggregate Data Tables**

Quantitative data collected via questionnaires to school superintendents:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Professional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development Plans (PDP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initiated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of remediation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plans initiated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dismissed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What year the district</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>implemented PDPs?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantitative data from Illinois State Board of Education:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Educators employed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student enrollment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District Type (Unit,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary, Secondary)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District size (small,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium, large)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E

Superintendent Questionnaire

CONSENT FORM

You are invited to participate in a study of teacher evaluation in Illinois. The purpose of this study is to explore how micropolitics influences qualified evaluators when identifying underperforming teachers. This research is being conducted by Mr. David Conrad, Doctor of Education degree student in Education Policy, Organization and Leadership at the University of Illinois at Urbana-Champaign and Professor Donald Hackmann.

Your participation in this research study is completely voluntary. Your decision to participate or not to participate will not affect your relationship in any way with your relationship with the University of Illinois. You may elect to terminate this activity at any time you begin to feel uncomfortable about the experience. Should you consent, you will participate in one survey, which should take about 10 minutes to fill-out. However, it may take much longer to collect the data needed to answer the survey questions. Survey questions will collect data on the number of teachers who started remediation plans, Professional Development Plans, or were dismissed by the board of education. Follow-up interviews may also be conducted, lasting approximately 15 minutes. With your permission, interviews will be audio recorded for the purposes of data analysis and will be transcribed, with all identifying information removed to protect confidentiality of the participants. You would receive a copy of the transcript by email attachment to double-check the information, and you may be contacted by telephone or email for clarification of your interview responses.

Your survey responses will be kept confidential and secure, and the results of the interviews will only be reported in the aggregate. This study will correlate implementation of these practices to publicly available demographic factors. Publication may include the use of quotations from your interview in educational presentations, on websites, and in professional publications, but pseudonyms will be used for all quotations so your responses cannot be attributed to you. There is no direct benefit to agreeing to participate in this study for participants, but participation in the study involves minimal risk. Through identifying the micropolitical factors that influence teacher evaluations, it is intended that policymakers and will benefit through improvements in teacher evaluation policy in Illinois.

In general, we will not tell anyone any information about you. When this research is discussed or published, no one will know that you were in the study. However, laws and university rules might require us to disclose information about you. For example, if required by laws or University Policy, study information which identifies you and the consent form signed by you may be seen or copied by the following people or groups: a) The university committee and office that reviews and approves research studies, the Institutional Review Board (IRB) and Office for Protection of Research Subjects; and b) University and state auditors, and Departments of the university responsible for oversight of research.

If you feel you have not been treated according to the descriptions in this form, or if you have any questions about your rights as a research subject, including questions, concerns, complaints, or to offer input, you may call the University of Illinois Office for the Protection of Research Subjects (OPRS) at 217-333-2670 or e-mail OPRS at irb@illinois.edu. If you have questions or comments regarding this study, please email Don Hackmann (dghack@illinois.edu) or phone (217.333.0230). Alternatively, you may email David Conrad (dconr2@illinois.edu) or phone (815.370.0263).

By indicating your agreement on question 1 and completing the survey, you are providing your consent for participation in this study. The survey will be available from XXXX, 2017 to XXXX, 2017.
1. I have read and understand this project and indicate my willingness to voluntarily take part in this research study. I have been given the opportunity to copy or print of this consent form for my records.

☐ Yes, I understand and agree to participate.
☐ No, I do not consent to participate.

2. Please select your district alphabetically listed A-M by city. (The purpose of identifying your school from this list is to correlate your school's publicly available demographic and academic performance data. All of your responses will remain confidential and will not be reported individually or identifiably.)

☐

3. Please select your school (Alphabetically listed N-Z by city). (The purpose of identifying your school from this list is to correlate your school's publicly available demographic and academic performance data. All of your responses will remain confidential and will not be reported individually or identifiably.)

☐
This survey is asking about certificated teaching staff in your school district, including teachers, school psychologists, social workers, school counselors, and school librarians. Please exclude school administrators, such as superintendents, principals, assistant principals, and curriculum directors. Only include deans of students or other quasi-administrative positions when these positions are evaluated under the same conditions as certificated teaching staff. Exclude paraprofessionals, cooks, administrative assistants, and other non-certificated positions.

Please provide as much data as you are able to. For the purposes of this research, we prefer as much data as possible but will accept whatever data you can provide.

4. How many certificated teaching staff—including school counselors, school psychologists, nonteaching school speech and language pathologists, school nurses and school social workers—started a Professional Development Plan triggered by a “needs improvement” rating as required by Section 24A-5(h) of the School Code during these school years

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td></td>
</tr>
<tr>
<td>2009-2010</td>
<td></td>
</tr>
<tr>
<td>2010-2011</td>
<td></td>
</tr>
<tr>
<td>2011-2012</td>
<td></td>
</tr>
<tr>
<td>2012-2013</td>
<td></td>
</tr>
<tr>
<td>2013-2014</td>
<td></td>
</tr>
<tr>
<td>2014-2015</td>
<td></td>
</tr>
<tr>
<td>2015-2016</td>
<td></td>
</tr>
<tr>
<td>2016-2017</td>
<td></td>
</tr>
</tbody>
</table>
5. How many certificated teaching staff—including school counselors, school psychologists, nonteaching school speech and language pathologists, school nurses and school social workers—started a remediation plan triggered by an “unsatisfactory” rating during these school years:

<table>
<thead>
<tr>
<th>Year</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td></td>
</tr>
<tr>
<td>2009-2010</td>
<td></td>
</tr>
<tr>
<td>2010-2011</td>
<td></td>
</tr>
<tr>
<td>2011-2012</td>
<td></td>
</tr>
<tr>
<td>2012-2013</td>
<td></td>
</tr>
<tr>
<td>2013-2014</td>
<td></td>
</tr>
<tr>
<td>2014-2015</td>
<td></td>
</tr>
<tr>
<td>2015-2016</td>
<td></td>
</tr>
<tr>
<td>2016-2017</td>
<td></td>
</tr>
</tbody>
</table>
6. How many certificated teaching staff—including school counselors, school psychologists, nonteaching school speech and language pathologists, school nurses and school social workers—were dismissed by your board of education during these school years:

<table>
<thead>
<tr>
<th>Year</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td></td>
</tr>
<tr>
<td>2009-2010</td>
<td></td>
</tr>
<tr>
<td>2010-2011</td>
<td></td>
</tr>
<tr>
<td>2011-2012</td>
<td></td>
</tr>
<tr>
<td>2012-2013</td>
<td></td>
</tr>
<tr>
<td>2013-2014</td>
<td></td>
</tr>
<tr>
<td>2014-2015</td>
<td></td>
</tr>
<tr>
<td>2015-2016</td>
<td></td>
</tr>
<tr>
<td>2016-2017</td>
<td></td>
</tr>
</tbody>
</table>

7. During which school year did your district initiate the professional development plans triggered by a “needs improvement” rating as required by Section 24A-5(h) of the School Code following the passage of the Performance Evaluation Reform Act (PERA)?

- [ ] 2010-2011
- [ ] 2011-2012
- [ ] 2012-2013
- [ ] 2013-2014
- [ ] 2014-2015
- [ ] 2015-2016
- [ ] 2016-2017
- [ ] Not implemented
This study will include a follow-up interview with currently practicing qualified evaluators in Illinois public schools. These include principals, assistant principals, directors, department chairpersons, and others involved in the evaluation of teachers. The study seeks qualified evaluators who 1) have identified one or more underperforming teachers by implementing a remediation plan and/or Professional Development Plan; and 2) have evaluated teachers before and after the implementation of SB7 and PERA. Would you be willing to share the name and contact information for an evaluator(s) you would recommend for this study?

8. Please share the name(s) of an evaluator(s) you would recommend for this study. (optional)

9. Please share the email address(es) (or other contact information) of an evaluator(s) you would recommend for this study. (optional)
* 10. Please provide the name of the person completing this form

* 11. Please provide the e-mail address of the person completing this form.

* 12. Please provide the phone number of the person completing this form.

13. Is there anything else you’d like to share about the survey or the study? (optional)
Appendix F

Email Confirming Study Eligibility and Soliciting Participation

Hello! My name is David Conrad. I am the principal at Manteno Middle School in Manteno, Illinois. I am completing my Doctor of Education degree in Education Policy, Organization and Leadership at the University of Illinois at Urbana-Champaign.

I am conducting my dissertation research on teacher evaluation policy. My advisor, Dr. Donald Hackmann, is directing my study. The purpose of this study is to explore how micropolitics influences qualified evaluators when identifying underperforming teachers. I am contacting you because you were recommended by your superintendent as someone to interview for this study.

If you are selected for inclusion in the study, you will participate in one interview lasting approximately 45-60 minutes. Follow-up interviews may also be conducted, lasting approximately 15 minutes. The interviews may be audiotaped with your permission. In addition, you will be asked to share any unused forms or policies related to teacher evaluation in your district. All forms will be blank and will not contain any personally identifiable information.

To determine your eligibility for study participation, I need to confirm that you meet the following criteria:

- Must be a currently practicing qualified evaluator in Illinois public schools who worked before and after the implementation of PERA/SB7 as evaluators of certificated staff;
- Currently may hold any school position that evaluates certified staff—including principals, assistant principals, department chairpersons, directors, teachers, and others engaged in teacher evaluation;
- Must have identified one or more teachers for a remediation plan, a Professional Growth Plan (PGP), or dismissal during your evaluation experience in Illinois.

If you meet these criteria, and you are interested in participation in this study, please respond to me via email (dconr2@illinois.edu) or phone (XXX-XXX-XXXX). If you express a willingness to participate, an informed consent form will be delivered to you by email and we will set up a time for a brief 10-minute screening interview over the phone. The screening interview is designed to ensure you met the criteria necessary for participation in this study. If you have questions or comments regarding this study, please contact my dissertation advisor, Dr. Donald Hackmann (dghack@illinois.edu).

Thank you very much!!!

David Conrad
Doctoral Student in Educational Organization and Leadership
University of Illinois at Urbana-Champaign
phone XXX.XXX.XXXX | email dconr2@illinois.edu
Appendix G

Screening Interview via Telephone

Introduction and Purpose

Today, I am calling because you have agreed to participate in a study that seeks to investigate teacher evaluation policy in Illinois. As indicated on the informed consent form, I will be taking detailed notes of this interview and all personally identifiable information will be removed and replaced by pseudonyms. Should you wish to stop the interview at any time, you may do so.

Questions

1. How many years of have you worked as an evaluator of certificated staff in Illinois public schools total?

2. How many years have you served as the principal of your current school, including this year as one full year?

3. Did you work as an evaluator of certificated staff in Illinois public schools before and after the SB7 and PERA reforms?

4. In what position(s) have you worked when you evaluated certificated staff in Illinois public schools (e.g., principal, assistant principal, department chairperson, director, teachers, and others engaged in teacher evaluation)?

5. Have you identified one or more teachers for a remediation plan, a Professional Development Plan (PDP), or dismissal when you worked as an evaluator of certificated staff in Illinois public schools? Briefly explain.

6. With which ethnicity do you identify?
   a. Hispanic / Latino
   b. Not Hispanic / Not Latino
   c. Choose not to identify

7. With which race/s do you identify (choose one or more regardless of ethnicity status selected above)?
   a. American Indian or Alaska Native
   b. Asian
   c. Black or African-American
   d. Native Hawaiian or Other Pacific-Islander
   e. White
   f. Choose not to identify

8. With which gender do you identify?
   a. Male
   b. Female
c. Other
d. Choose not to identify

9. Is there anything else you would like for me to consider when determining your eligibility for the study?

This concludes the screening interview. Should you and your district be selected for participation in the study, I will notify you by email as soon as possible.
Appendix H

Informed Consent Documents

Qualitative Phase Consent

A Mixed Methods Study of Teacher Evaluation and Micropolitics in Illinois

CONSENT FORM

You are invited to participate in a study of teacher evaluation in Illinois. The purpose of this study is to explore how micropolitics influences qualified evaluators when identifying underperforming teachers. This research is being conducted by Mr. David Conrad, Doctor of Education degree student in Education Policy, Organization and Leadership at the University of Illinois at Urbana-Champaign and Professor Donald Hackmann.

Your participation in this research study is completely voluntary. Your decision to participate or not to participate will not affect your relationship in any way with your relationship with the University of Illinois. You may elect to terminate this activity if at any time you begin to feel uncomfortable about the experience. Should you consent, you will participate in one survey, which should take about 10 minutes to fill-out. However, it may take much longer to collect the data needed to answer the survey questions. Survey questions will collect data on the number of teachers who started remediation plans, Professional Development Plans, or were dismissed by the board of education. You may be contacted by e-mail or by telephone to follow-up on your responses, for the purposes of clarifying any questions and ensuring the accuracy of your survey responses.

Your survey responses will be kept confidential and secure, and the results of the interviews will only be reported in the aggregate. This study will correlate implementation of these practices to publicly available demographic factors. Publication may include the use of quotations from your interview in educational presentations, on websites, and in professional publications, but pseudonyms will be used for all quotations so your responses cannot be attributed to you. There is no direct benefit to agreeing to participate in this study for participants, but participation in the study involves minimal risk. Through identifying the micropolitical factors that influence teacher evaluations, it is intended that policymakers and will benefit through improvements in teacher evaluation policy in Illinois.

In general, we will not tell anyone any information about you. When this research is discussed or published, no one will know that you were in the study. However, laws and university rules might require us to disclose information about you. For example, if required by laws or University Policy, study information which identifies you and the consent form signed by you may be seen or copied by the following people or groups: a) The university committee and office that reviews and approves research studies, the Institutional Review Board (IRB) and Office for Protection of Research Subjects; and b) University and state auditors, and Departments of the university responsible for oversight of research.

If you feel you have not been treated according to the descriptions in this form, or if you have any questions about your rights as a research subject, including questions, concerns, complaints, or to offer input, you may call the University of Illinois Office for the Protection of Research Subjects (OPRS) at 217-333-2670 or e-mail OPRS at irb@illinois.edu. If you have questions or comments regarding this study, please email Don Hackmann (dghack@illinois.edu) or phone (217.333.0230). Alternatively, you may email David Conrad (dconrad2@illinois.edu) or phone (815.370.0263).

By indicating your agreement on question 1 and completing the survey, you are providing your consent for participation in this study. The survey will be available from XXXX, 2017 to XXXX, 2017.

* 1. I have read and understand this project and indicate my willingness to voluntarily take part in this research study. I have been given the opportunity to copy or print this consent form for my records.

   □ Yes, I understand and agree to participate.
   □ No, I do not consent to participate.

University of Illinois at Urbana-Champaign
Institutional Review Board

Approved: 9-23-2017
IRB #: 17-0045
You are invited to participate in a study of teacher evaluation in Illinois. The purpose of this study is to explore how micropolitics influences qualified evaluators when identifying underperforming teachers. This research is being conducted by Mr. David Conrad, Doctor of Education degree student in Education Policy, Organization and Leadership at the University of Illinois at Urbana-Champaign and Professor Donald Hackmann.

Your participation in this research study is completely voluntary. Your decision to participate or not to participate will not affect your relationship in any way with your relationship with the University of Illinois. You may elect to terminate this activity if at any time you begin to feel uncomfortable about the experience. Should you consent, you will participate in one interview, which should last no longer than 1 hour. Interview questions will focus on how micropolitics influences teacher evaluation. Follow-up interviews may also be conducted, lasting approximately 15 minutes. With your permission, interviews will be audio recorded for the purposes of data analysis and will be transcribed, with all identifying information removed to protect confidentiality of the participants. You also may choose to voluntarily provide documents or other artifacts to assist the researchers in understanding the school district’s implementation of teacher evaluation policy. Please note that voluntarily provided documents of student records will not be accepted; this study does not require access to any student records. You will receive a copy of the transcript by email attachment to double-check the information, and you may be contacted by telephone or email for clarification of your interview responses.

Your interview responses will be kept confidential and secure, and the results of the interviews will only be reported in the aggregate. Publication may include the use of quotations from your interview in educational presentations, on websites, and in professional publications, but pseudonyms will be used for all quotations so your responses cannot be attributed to you. There is no direct benefit to agreeing to participate in this study for participants, but participation in the study involves minimal risk. Through identifying the micropolitical factors that influence teacher evaluations, it is intended that policymakers and will benefit through improvements in teacher evaluation policy in Illinois.

In general, we will not tell anyone any information about you. When this research is discussed or published, no one will know that you were in the study. However, laws and university rules might require us to disclose information about you. For example, if required by laws or University Policy, study information which identifies you and the consent form signed by you may be seen or copied by the following people or groups: a) The university committee and office that reviews and approves research studies, the Institutional Review Board (IRB) and Office for Protection of Research Subjects; and b) University and state auditors, and Departments of the university responsible for oversight of research.

If you feel you have not been treated according to the descriptions in this form, or if you have any questions about your rights as a research subject, including questions, concerns, complaints, or to offer input, you may call the University of Illinois Office for the Protection of Research Subjects (OPRS) at 217-333-2670 or e-mail OPRS at irb@illinois.edu. If you have questions or comments regarding this study, please email Don Hackmann (dghack@illinois.edu) or phone (217.333.0230). Alternatively, you may email David Conrad (dcon2@illinois.edu).

________________________
I have read and understand this project and indicate my willingness to voluntarily take part in this research study. I have been given a copy of this consent form for my records.

I consent to be interviewed for this study and to have my interview audio recorded and transcribed. [ ] Yes [ ] No (check one)
I consent to be to be re-contacted for follow-up interviews [ ] Yes [ ] No (check one)

Printed Name: __________________________________________ Email: __________________________________________
Signature: __________________________________________ Date: __________________________________________

University of Illinois at Urbana-Champaign
Institutional Review Board

Approved: 6-27-2017

IRB #: 17845
Appendix I

Semi-Structured Interview Protocol

Introduction and Purpose

Today, we are meeting because you have agreed to participate in a study that seeks to investigate teacher evaluations in Illinois. As indicated on the informed consent form, I will be recording this interview and all personally identifiable information will be removed and replaced by pseudonyms. Should you wish to stop the interview at any time, you may do so.

I will start by asking you some questions about your school.

1. In your opinion, what percentage of certificated staff in your school district are underperforming?
2. Now I will share the data your district provided to me in a previous phase of the study. Please share your reactions to the data.

I will now ask you about the ground rules and procedures for teacher evaluations (Bridges and Groves, 1999).

3. How have your teacher evaluations changed since PERA / SB7? Have the rules or procedures in your current teacher evaluation plan made it easier or harder to issue an underperforming rating than prior to PERA / SB7? Please explain.
4. Have you ever had an underperforming teacher receive a Reduction in Force notice? Please explain.
5. What support would you need to improve the identification of underperforming teachers?

Next, I will ask you about persons or interests that have predominated or been frustrated by PERA / SB7 (Bridges and Groves, 1999).

6. How have teachers responded to the changes in teacher evaluation since PERA / SB7? Has teaching and learning improved? Please explain.
7. What experiences have you had with remediation plans or Professional Development Plans? Have they improved teachers? Please explain. (Potential follow-up: knowing your experience, would you do it again?)

Next, I will ask you about the sources and levels of power, coalitions, and strategies of various actors (Bridges and Groves, 1999).

8. Has identifying underperforming teachers had any negative effects or repercussions? How did this effect the relationships you had with other teachers?
9. Have you ever been in a situation when you were unable or unwilling to give a low rating to an underperforming teacher? Please explain.
10. Has anyone (such as a board of education member, supervisor, teacher, or the teacher union) ever tried to influence your decision to evaluate someone with a higher or lower summative rating than you felt was earned? Please explain.

Next, I will ask about excesses or abuses (Bridges and Groves, 1999)

11. Have teachers been evaluated fairly in your district? Have any changes to your evaluation process been made because of fairness or other issues?

Next, we will talk the balance of interests of parents and students having quality teachers with the interests of employees in job security and fair treatment (Bridges and Groves, 1999).

12. Has pressure from students or parents influenced your evaluation ratings?

Finally, we will conclude our interview with a few wrap-up questions.

13. Do you have any other issues or information you would like to share, that I haven’t asked you, that you want to ensure that I collect for this study?
14. Is there any written documentation your wish to share?

Thank you for your participation. I may be contacting for follow-up questions later and to give you the opportunity to review my findings regarding the study. Thank you!
Appendix J

Institutional Review Board Approval

**Exempt Approval - IRB #17845**
Institutional Review Board

- **Sent:** Tuesday, June 27, 2017 3:09 PM
- **To:** Hackmann, Donald
- **Cc:** Conrad, David L
- **Attachments:** 17845 RTA.PDF (109 KB) ; 17845 Consent.pdf (77 KB)

**IRB EXEMPT APPROVAL.**

**RPI Name:** Dr. Donald Hackman  
**Project Title:** *A Mixed Methods Study of Teacher Evaluation Reforms and Micropolitics in Illinois*  
**IRB #:** 17845  
**Approval Date:** June 27, 2017

Thank you for submitting the completed IRB application form and related materials. Your application was reviewed by the UIUC Office for the Protection of Research Subjects (OPRS). OPRS has determined that the research activities described in this application meet the criteria for exemption at 45CFR46.101(b)(2). This message serves to supply OPRS approval for your IRB application.

Please contact OPRS if you plan to modify your project (change procedures, populations, consent letters, etc.). Otherwise you may conduct the human subjects research as approved for a period of five years. Exempt protocols will be closed and archived at the time of expiration. Researchers will be required to contact our office if the study will continue beyond five years.

Copies of the attached, date-stamped consent form(s) are to be used when obtaining informed consent.

We appreciate your conscientious adherence to the requirements of human subjects research. If you have any questions about the IRB process, or if you need assistance at any time, please feel free to contact me at OPRS, or visit our website at http://oprs.research.illinois.edu

Sincerely,

Jennifer Ford, MS  
Human Subjects Research Specialist, Office for the Protection of Research Subjects

Attachment(s): Research team attachment, consent form

C: David Conrad
Dear Dr. Hackman & Mr. Conrad,

This message serves to supply UIUC IRB approval for the minor modifications being made to your protocol IRB #17845, *A Mixed Methods Study of Teacher Evaluation Reforms and Micropolitics in Illinois*. This amendment approves the following changes:

- Inclusion of a phase 1 online survey

It has been determined that the research activities described in this application still meet the criteria for exemption at 45CFR46.101(b)(1, 2).

This determination of exemption only applies to the research study as submitted. Please note that additional modifications to your project need to be submitted to the IRB for review and exemption determination or approval before the modifications are initiated.

We appreciate your conscientious adherence to the requirements of human subjects research. If you have any questions about the IRB process, or if you need assistance at any time, please feel free to contact me at the OPRS office, or visit our website at [http://oprs.research.illinois.edu](http://oprs.research.illinois.edu).

Jennifer N. Ford, MS  
Human Subjects Research Specialist | Office for the Protection of Research Subjects  
University of Illinois Urbana-Champaign  
805 West Pennsylvania Avenue | Urbana, IL 61801  
Direct: (217) 300-2022 | Fax: (217) 333.0405 | Email: jford2@illinois.edu
## Appendix K

### Interview Log

<table>
<thead>
<tr>
<th>Participant</th>
<th>School District</th>
<th>Date</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aidan O’Brien</td>
<td>Irving</td>
<td>December 21, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Amanda Ashbee</td>
<td>Sunfields</td>
<td>December 18, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Ann Keaton</td>
<td>Napa River</td>
<td>December 20, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Charlie Ramirez</td>
<td>Churchill</td>
<td>December 18, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Corey Hammer</td>
<td>Ferndale</td>
<td>December 6, 2017</td>
<td>In Person</td>
</tr>
<tr>
<td>Diana Legend</td>
<td>Rosedale</td>
<td>December 6, 2017</td>
<td>In Person</td>
</tr>
<tr>
<td>Dustin Ross</td>
<td>St. Paul</td>
<td>December 18, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Emma Moore</td>
<td>Gregson</td>
<td>January 16, 2018</td>
<td>Telephone</td>
</tr>
<tr>
<td>Eric Graves</td>
<td>Founders</td>
<td>December 18, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Frank Steele</td>
<td>Bayview</td>
<td>December 15, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Glen Tucker</td>
<td>Crater Bay</td>
<td>December 15, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Harold Cooper</td>
<td>Kappel</td>
<td>January 9, 2018</td>
<td>Telephone</td>
</tr>
<tr>
<td>Jeffrey Shannon</td>
<td>Highlane</td>
<td>January 5, 2018</td>
<td>Telephone</td>
</tr>
<tr>
<td>Jordyn Harris</td>
<td>Hillman</td>
<td>January 12, 2018</td>
<td>Telephone</td>
</tr>
<tr>
<td>Melody Roberts</td>
<td>Pittman</td>
<td>December 8, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Michelle Hale</td>
<td>Austin</td>
<td>December 15, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Molly Martin</td>
<td>Logan</td>
<td>December 18, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Rebecca Silver</td>
<td>Jasper</td>
<td>January 12, 2018</td>
<td>Telephone</td>
</tr>
<tr>
<td>Stephen Pierce</td>
<td>Maple Ridge</td>
<td>December 18, 2017</td>
<td>Telephone</td>
</tr>
<tr>
<td>Vickie Wells</td>
<td>Abbidale</td>
<td>December 20, 2017</td>
<td>Telephone</td>
</tr>
</tbody>
</table>