



2014

## The Perceived Impact of Teacher Performance Ratings on the Teacher Evaluation Process: Voices from the Field

Brian B. Bullis  
*Loyola University Chicago*

Follow this and additional works at: [https://ecommons.luc.edu/luc\\_diss](https://ecommons.luc.edu/luc_diss)



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

---

### Recommended Citation

Bullis, Brian B., "The Perceived Impact of Teacher Performance Ratings on the Teacher Evaluation Process: Voices from the Field" (2014). *Dissertations*. 889.  
[https://ecommons.luc.edu/luc\\_diss/889](https://ecommons.luc.edu/luc_diss/889)

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).  
Copyright © 2013 Brian B. Bullis

LOYOLA UNIVERSITY CHICAGO

THE PERCEIVED IMPACT OF TEACHER PERFORMANCE RATINGS ON THE  
TEACHER EVALUATION PROCESS: VOICES FROM THE FIELD

A DISSERTATION SUBMITTED TO  
THE FACULTY OF THE GRADUATE SCHOOL OF EDUCATION  
IN CANDIDACY FOR THE DEGREE OF  
DOCTOR OF EDUCATION

PROGRAM IN ADMINISTRATION AND SUPERVISION

BY

BRIAN BULLIS

CHICAGO, ILLINOIS

MAY 2014

Copyright by Brian Bullis, 2014

All rights reserved.

## ACKNOWLEDGEMENTS

A significant undertaking such as what is found on the following pages, and also what these pages signify in the form of a doctoral degree in Administration and Supervision from Loyola University Chicago, does not come together without the support and guidance of an incredible cast of characters.

First, thank you to my family. To my parents, Paul and Bridget, for impressing upon me the value of an education, a love for learning from the earliest age, and for always instilling confidence in me. To my loving wife Ann for her support, patience, wisdom, and shared love for the field of education. To my daughters, Catie and Ellie, for being constant reminders that my actions today will play a role in shaping who they are in the future. To Keegan, Meghan, and Quinn for being remarkable siblings who always show unconditional support for our family while pushing me to want to be someone for which they would be proud. To Rich and Nancy for their support and appreciation of this doctoral process and in helping out, in little and in big ways, to allow me to get there.

Thank you next to my Loyola educational family. To Dr. Marla Israel, my Dissertation Chair, for reminding me that if I put my absolute trust in someone great things can happen. Her guidance, support, honesty, intelligence, and positivity were critical and appreciated beyond comprehension throughout this process. To Dr. Janis Fine and Dr. Susan Sostak for providing a critical ear and a true investment in my success throughout this doctoral program and the dissertation; including their willingness to be on

my dissertation committee. To all of the other knowledgeable and passionate Loyola professors that contributed their respective expertise to this doctoral experience.

Thank you to my professional education family. To the Deerfield doctoral cohort for being a constant source of support, family, humor, mediocre food, incredible intellect, and a shared desire to grow and be lifelong learners. To the Deerfield Public Schools for supporting this endeavor and valuing the growth and development of their leaders; I am forever indebted. To every student and educator that I have encountered to this point in my career for collectively shaping my love for this profession, love for learning, and desire to give back.

Lastly, thank you to the hundreds of principals in Florida and Massachusetts that were kind enough to take valuable time out of their own personal and professional lives to participate in this survey. Due to the kindness of others this dissertation was made possible.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
ABSTRACT.....	xi
CHAPTER	
I. INTRODUCTION .....	1
Purpose of the Study .....	1
Research Questions .....	15
Significance of the Study to the Field of Educational Leadership.....	17
Proposed Methodology .....	17
Areas of Related Literature .....	19
Conceptual Framework .....	21
Summary .....	24
II. LITERATURE REVIEW .....	25
Introduction.....	25
A Brief Historical Overview of Teacher Evaluation .....	27
Teacher Evaluation and Teacher Recognition .....	30
Teacher Evaluation and Teacher Effectiveness .....	33
Teacher Evaluation and Teacher Growth .....	36
Teacher Evaluation and Teacher Remediation .....	39
Teacher Evaluation and Teacher Dismissal.....	41
The Relationship between Principal and Teacher.....	45
Courageous and Strategic Conversations.....	50
Teacher Evaluator Impact.....	52
Culture and Climate .....	55
Promoting Social Justice.....	57
Current Trends and Impacts on Teacher Evaluation.....	59
Race to the Top .....	61
Charlotte Danielson’s Framework for Teaching.....	62
Teacher Evaluation Reform in Illinois.....	67
Teacher Evaluation Reform in Massachusetts .....	74
Teacher Evaluation Reform in Florida .....	79
Summary .....	83

III. METHODOLOGY .....	84
Introduction.....	84
Research Design.....	86
Participants.....	88
Procedures for Data Collection.....	90
Data Analysis .....	94
Ethical Considerations .....	97
Minimization of Bias .....	98
Validity and Reliability.....	99
Limitations .....	100
Summary .....	101
IV. PRESENTATION OF DATA .....	103
Introduction.....	103
Review of the Survey Administration.....	104
Data Presentation .....	108
Demographic Information.....	108
Evaluation Background, Experience, and Perspective .....	111
Teacher Recognition .....	118
Teacher Effectiveness .....	123
Teacher Growth .....	129
Teacher Remediation .....	137
Teacher Dismissal.....	146
Additional Comments .....	152
Conceptual Framework.....	153
Summary .....	154
V. ANALYSIS OF DATA.....	157
Introduction.....	157
Conclusions.....	158
Research Question 1 .....	158
Research Question 2 .....	160
Research Question 3 .....	165
Research Question 4 .....	167
Research Question 5 .....	168
Instruction as the Most Impacted Domain.....	169
Professional Responsibilities as Least Impacted Domain .....	171
Need for More Time .....	173
More Teachers on Remediation.....	175
Less Teachers Being Dismissed.....	176
Analysis of Teacher Rating Protests and Rating Interference with Growth .....	177
Teacher Morale and Stress are Recurring Unintended Themes.....	179

Limitations of the Study.....	180
Recommendations for Future Research.....	182
Significance to Educational Leadership Practice and Preparation .....	184
Time Element.....	184
Principal as the Instructional Lever .....	186
Summary of Findings.....	188

## APPENDIX

A. CONSENT LETTER .....	192
B. SURVEY INSTRUMENT.....	196
C. FOLLOW-UP LETTER #1.....	202
D. FOLLOW-UP LETTER #2.....	204

REFERENCES .....	206
------------------	-----

VITA.....	213
-----------	-----

## LIST OF TABLES

Table	Page
1. State Proficiency Levels .....	11
2. Sergioivanni's Sources of Authority .....	49
3. Teacher Expertise and Student Achievement.....	58
4. Danielson Levels of Performance Descriptors .....	66
5. Illinois Poor Teacher Evaluation Consequences .....	72
6. Massachusetts Teacher Rating Definitions .....	75
7. Massachusetts Teacher Evaluation Cycle .....	78
8. Florida Poor Teacher Evaluation Consequences.....	81
9. Comparison of Illinois, Massachusetts, and Florida Teacher Evaluation Systems.....	82
10. Data Analysis Framework under Four-tier Teacher Evaluation System.....	96
11. Data Analysis Framework under Four-Tier Teacher Evaluation System for All Responses .....	154
12. Data Analysis Framework under Four-Tier Teacher Evaluation System for Target Population .....	155
13. Summary of the Intended Impacts of the Four-Tier Rating System .....	159
14. Summary of the Unintended Impacts of the Four-Tier Rating System.....	161
15. Data Analysis Framework under Four-Tier Teacher Evaluation System for All Responses .....	170
16. Data Analysis Framework under Four-Tier Teacher Evaluation System for Target Population .....	171

## LIST OF FIGURES

Figure	Page
1. Charlotte Danielson’s Domains and Elements of the Framework for Teaching.....	64
2. Massachusetts Teacher Rating Chart .....	75
3. Data Analysis Procedures.....	95
4. Identification of the Survey Target Population .....	107
5. State that Public School Principal Evaluated Teachers in the Past Year .....	109
6. Educational Background .....	110
7. Current Age of Respondent.....	111
8. Years of Teaching Experience .....	112
9. Years of Principal Experience .....	113
10. Approximate Number of Teacher Evaluated Per Year by Principal .....	114
11. Total Years Principal has Evaluated Teachers Using a Four-Tier Performance Rating System .....	115
12. Years Evaluating Teachers in Principals’ Current State .....	116
13. Comparison of the Four-Tier Rating System to Principals’ Prior Rating System .....	117
14. Number of Performance Ratings Preferred for Teacher Evaluation .....	118
15. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Recognize Teacher Excellence.....	120
16. The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Recognition .....	121
17. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Recognition .....	122
18. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Recognize Teacher Effectiveness.....	124
19. Necessity for a System with More Than One Tier to Recognize Teacher .....	125
20. Necessity to Have a System with More than One Tier to Recognize Teacher Attainment of Standards (i.e., “Proficient” and “Excellent”).....	126
21. The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Effectiveness .....	127
22. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Effectiveness .....	129
23. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Promote Teacher Growth .....	130
24. Impact of the Performance Rating on Teacher Growth .....	131

25. Performance Rating Promotion of Identified Areas for Growth in the Evaluation.....	132
26. Frequency of Teacher Challenges or Protests of Assigned Performance Rating Under the Four-Tier System .....	133
27. Frequency of Teacher Challenges or Protests of Assigned Performance Rating Under the Principals' Previous Evaluation System .....	134
28. The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Growth.....	135
29. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Growth.....	137
30. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Identify Teacher for Remediation .....	138
31. Teachers per Year Put on Remediation under the Current Four-Tier Rating System .....	139
32. Teachers per Year Put on Remediation under the Previous Evaluation System .....	140
33. Necessity for a System with More Than One Tier to Identify Standard Deficiency .....	141
34. Necessity to Have a System with More than One Tier to Recognize Teacher Deficiency in Meeting Standards (i.e., "Needs Improvement" and "Unsatisfactory").....	142
35. The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Remediation.....	144
36. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Remediation.....	145
37. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Identify Teacher for Dismissal .....	147
38. Teachers per Year Terminated under the Current Four-Tier Rating System .....	148
39. Teachers per Year Dismissed under the Previous Evaluation System.....	149
40. The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Dismissal .....	150
41. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Dismissal .....	151
42. Necessity for a System with More Than One Tier to Recognize Teacher Standard Attainment.....	166
43. Necessity for a System with More Than One Tier to Identify Standard Deficiency .....	166
44. Summary of Study Findings.....	191

## ABSTRACT

The purpose of this study was focused on the various perceived impacts created by the expansion to a four-tier teacher performance evaluation rating model which would inform educational leaders in the State of Illinois. By studying the experiences of principals in two other states who previously underwent the same change, Florida and Massachusetts, a number of insights were found that can serve to inform Illinois.

The intended impacts found from expanding the performance ratings included the promotion of teacher growth, recognition of teacher excellence, promotion of teacher remediation, and support in dismissing ineffective teachers. The unintended impacts that were found included low teacher morale, interference with teacher growth, teacher stress, and difficulty dismissing teachers; while others found no unintended impact.

In regards to the intended impact of having multiple tiers for standard attainment or deficiency the research found that these tiers help to delineate the performance of those meeting standards and those not meeting standards while no significant unintended impacts were found.

The most significant of the messages to inform Illinois included the fact that instruction was the most positively impacted of the Charlotte Danielson domains while professional responsibilities was the least impacted. Also, it was realized that more time was needed both within and across academic years to more effectively meet the demands of the evaluation process.

Given these findings the researcher posited that two major lessons learned. First, system reform needs to be given time in order to be implemented effectively and yield the desired results. Second, principals must dedicate time and energy to serving as the instructional lever in an educational organization and the school will improve under the expanded teacher evaluation rating system.

## CHAPTER I

### INTRODUCTION

#### **Purpose of the Study**

The purpose of this study is to explore and measure the principals' perception of the impact of teacher performance ratings on recognition, effectiveness, growth, remediation, and dismissal. In light of recent legislation in Illinois, including the Performance and Evaluation Reform Act of 2010 and the Senate Bill 7 Public Act 97-0008 in 2011, the teacher evaluation process in Illinois has undergone major changes in its utilization and significance. One specific change that was mandated beginning on September 1, 2012 was that all teachers would be evaluated using a four-tier performance rating system in Illinois. The four evaluation performance ratings are now "excellent," "proficient," "needs improvement," and "unsatisfactory."

The State of Illinois is waiting to see what kind of impact the new expanded teacher rating system has on teacher performance and subsequent student outcomes. School administrators using similar models in the past may have encountered a variety of responses from teachers in regards to the performance rating they received and may have found varied levels of effectiveness in the implementation of this model impacting teacher growth and perceived effectiveness. The intent of this study is to inform principals regarding the perceived impacts of the expanded four-tier performance rating system in Illinois in regards to teacher recognition, teacher effectiveness, and teacher

growth, as well as remediating and dismissing ineffective teachers. This research was done by studying two other states, Florida and Massachusetts, that have recently undergone a similar teacher evaluation transformation which will inform educational leaders in Illinois. Their data was analyzed to determine if their perceptions possess any meaningful information that will help to answer the overarching research questions of this study.

The research questions focus on the various perceived impacts, intended and unintended, created by a state mandated expansion to a four-tier teacher performance evaluation rating model. In order to capture these perceived impacts, the following research questions were researched and answered:

From the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent”, “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?

- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

The *Race to the Top* Program request for proposals was released in November of 2009 as part of the American Recovery and Reinvestment Act of 2009 (ARRA). ARRA provided \$4.35 billion for this competitive grant program. States were invited to apply for funds under this grant and were rewarded for implementing innovative strategies which would lead to improved student performance (“Race to the Top,” 2009).

One criterion under the grant requirements, according to the *Race to the Top* Executive Summary (2009), was “Improving teacher and principal effectiveness based on performance.” Specifically, points were awarded to applicants who “differentiate effectiveness using multiple rating categories that take into account data on student growth...as a significant factor.” The application did not elaborate on specifically how many performance rating categories were to be used or what descriptors should be used to identify the ratings.

Illinois applied three times for *Race to the Top* funds and was denied the first two times with the third application being accepted in December of 2011. During Phase One, which was submitted in January of 2010, Illinois cited their commitment to robust teacher

evaluations through the enactment of the Performance Evaluation Reform Act (PERA) of 2010. The State of Illinois *Race to the Top* Application for Initial Funding (2010) stated at the beginning of the section on performance evaluation systems that “Teacher and principal evaluation in Illinois is broken” (p. 94). It goes on to state that in three of the state’s largest districts they found that 92.6% of teacher were rated “superior” or “excellent,” 7% were rated as “satisfactory,” and 0.4% were rated “unsatisfactory.”

The grant proposal explained the expansion from three rating categories for teachers to four rating categories with the addition of a “Needs Improvement” category which was added to the categories of “Excellent,” “Proficient,” and “Unsatisfactory” beginning in 2012-13. PERA also eliminated the prior ability of school districts to obtain waivers to bypass this rating system which over 60 schools had obtained in the past to often implement a binary rating system (Illinois State Board of Education, 2010).

The State of Illinois *Race to the Top* Phase Two application was submitted in June of 2010 after PERA was approved by the Illinois General Assembly in January 2010. In addition to reiterating the information and proposed action steps from Phase One the application also presented specific detail regarding the role of the Performance Evaluation Advisory Council (PEAC) in implementing the PERA initiatives. It is stated that PEAC membership would be comprised of “practicing teachers and principals, as well as the statewide associations representing them” (Illinois State Board of Education, 2010). Under PERA, the PEAC was expected to meet quarterly up through June 30, 2017 although it was meeting on a much more regular basis. One of the tasks assigned to PEAC, which was detailed in Phase Two, was to begin the process of defining state standards of evaluation feedback that were both timely and constructive. Constructive

was specifically identified as feedback that “must define specific areas for improvement and actionable goals in order for a teacher or principal to achieve the next highest evaluation rating” (“State of Illinois Phase Two,” 2010).

Illinois submitted Phase Three of the *Race to the Top* in December of 2011. Little had changed in regards to teacher performance ratings in this latest version (“State of Illinois Phase Three,” 2011). On December 22, 2011, Illinois was officially notified that it received funding from *Race to the Top* Phase Three. Funds were awarded in the amount of \$42,818,707 (“Race to the Top: Phase 3 Award Letter, Illinois, 2011). Thirty-five school districts in Illinois agreed to be part of this grant, including the Chicago Public School district which was awarded just over \$19,000,000 of the total grant award (“Illinois Race to the Top Phase 3: Allocations for Participating LEAs,” 2012).

The Performance Evaluation Reform Act was approved by Governor Pat Quinn on January 15, 2010. Beyond the Act’s function of expanding the teacher performance rating categories, it required student growth to be a “significant” factor in teacher and principal performance categories. This growth function was implemented for all principals statewide and for teachers in 300 Chicago Public Schools beginning with the 2012-13 school year. The following year the remaining CPS schools integrated student growth measures into teacher evaluations. Beginning in 2015-16 the lowest-performing 20% of school districts will incorporate student growth measures in their teacher evaluations followed by the remaining Illinois school districts in 2016-2017 (Performance Evaluation Reform Act [PERA], 2010).

Section 5 of the Performance Evaluation Reform Act began with a focus on the State’s findings. These included the concept that effective teachers and effective school

leaders play a major role in student achievement. The findings focused on how the Illinois school district performance evaluation systems “fail to adequately distinguish between effective and ineffective teachers and principals” (PERA, 2010). The findings continued by citing that in a recent study of the three largest Illinois school districts that “out of 41,174 teacher evaluations performed over a 5-year period, 92.6% of teachers were rated “superior” or “excellent,” 7% were rated “satisfactory,” and only 0.4% were rated “unsatisfactory”” (PERA, 2010). No comments were made regarding what an appropriate ratio should be or if these statistics represented the majority of Illinois school districts. However, these numbers did provide a frame of reference for the State to reflect upon.

Along with a revision to the performance ratings used in Illinois, PERA also stated that the performance evaluation system must go beyond measuring professional competencies and must also assess student growth. In doing so it was under the direction of PERA that the State of Illinois and individual school districts “must ensure that performance evaluation systems are valid and reliable and contribute to the development of staff and improved student achievement outcomes” (PERA, 2010). In order to establish student growth measures as a component of teacher performance ratings each district was directed to create a joint committee comprised of an equal representation of teachers and administrators as selected by the district and its teachers. If this group could not reach an agreement on the plan then the district would adopt the model evaluation plan selected by the State.

In regards to performance ratings, it was stated in Section 24A-5 that “each teacher in contractual continued service is evaluated at least once in the course of every 2

school years and probationary teachers must be evaluated annually. However, any tenured teacher...whose performance is rated as either “needs improvement” or “unsatisfactory” must be evaluated at least once in the school year following the receipt of such rating” (PERA, 2010). Those teachers that are given a rating of “needs improvement” will begin the process of a professional development plan with a focus on the areas that need to improve along with district supports in these identified areas. Those receiving an “unsatisfactory” will begin the process of a remediation plan in the event that the deficiencies are deemed ‘remediable.’ PERA states that the remediation plan for unsatisfactory, tenured teachers for all school districts, “shall provide for 90 school days of remediation within the classroom” (PERA, 2010). If a teacher with a “needs improvement” or “unsatisfactory” is able to achieve a rating equal to or better than “satisfactory” he/she is to be reinstated in the district’s regular evaluation schedule cycle along with all other “satisfactory” and “excellent” teachers in the teacher’s respective tenured or non-tenured track. If a teacher with a “needs improvement” or “unsatisfactory” fails to complete any part of his/her remediation plan with a rating of “satisfactory” or better he/she is to be dismissed in accordance with Section 24-12 or 34-85 of the School Code.

Senate Bill 7 Public Act 97-0008, also known as Ed Reform, elaborated further on how performance evaluation categories impact teachers, school districts, and the State. Section 24-1.5 further emphasized the new importance of the performance ratings. Relevant experience will not be considered as a factor in filling a vacant teaching position unless other factors (certifications, qualifications, merit and ability – including performance evaluations) are equal.

Senate Bill 7 also incorporated the significance of the performance evaluation ratings into the process for teachers to obtain tenure. Probationary teachers can obtain tenure in one of three ways. The first path to tenure is if the teacher obtains a rating of at least “proficient” in their fourth school term in addition to at least a “proficient” in the second or third school term.” A teacher that fails to meet these requirements is mandated by law to be dismissed at the end of the fourth school term. The second path to tenure is if the teacher earns three consecutive terms of “excellent.” The third possible path to tenure is if the teacher had previously achieved tenure in a different district and received at least a “proficient” for his/her two most recent post-PERA evaluations followed by a rating of “excellent” for the first two school terms in his/her new district. Probationary teachers may still be non-renewed or dismissed by school boards with certain provisions. These provisions include the stipulation that no reasons are required for a first or second-year teacher, but a third-year teacher must be given a reason if he/she has received three “excellent” ratings and a fourth-year teacher must be given reasons unless he/she cannot acquire tenure due to poor performance ratings (Illinois Pension Code Senate Bill 7, 2010).

Reductions-in-force (RIFs) are another major component of Senate Bill 7 that incorporate performance evaluation categories into the decision making process. All teachers are now placed into one of four groups for every position they are qualified to teach based on performance evaluation categories. When RIFs do occur the dismissals will begin with Group 1 and move toward Group 4 sequentially from there. To simplify a more elaborate process Group 1 consists of probationary teachers who have not yet been evaluated while Group 2 consists of teachers with either a “needs improvement” or

“unsatisfactory” performance rating on either of their last two ratings. Group 3 consists of teachers who received a “proficient” rating on both of their last two evaluations and Group 4 consists of teachers who either received “excellent” on their last two performance evaluations or received “excellent” ratings on two of their last three evaluations and a third rating of “proficient” during that span. Length of teacher service only plays a role as a tiebreaker within the different groups and does not supersede the groups (Illinois Pension Code Senate Bill 7, 2010).

Section 21-23(a) of Senate Bill 7 addresses suspension or revocation of certificates and has specific language regarding unsatisfactory ratings. It defines incompetency as “two or more school terms of service for which the certificate holder has received an unsatisfactory rating on a performance evaluation....within a period of 7 school terms of service” (Illinois Pension Code Senate Bill 7, 2010, p. 8). At this point the decision on whether or not to take action against a teacher’s certificate lies within the purview of the State Superintendent. If a hearing is needed it will take place in the educator’s educational service region “in accordance with rules adopted by the State Board of Education, in consultation with the State Teacher Certification Board” (Illinois Pension Code Senate Bill 7, 2010, p. 10). Any decision made by the State Certification Board is considered a final administrative decision.

The Performance Evaluation Advisory Council (PEAC) was formed in conjunction with preparation for the Illinois Phase 2 application for *Race to the Top* and was charged with the responsibility to lead all of the state evaluation efforts (“State of Illinois Phase Two,” 2010). The group began to collaborate in the fall of 2009 to improve the principal evaluation process. PEAC was charged with the task of developing

a state model for teacher and principal evaluation under PERA. The members of PEAC stated in a June 1, 2011 presentation that there was a need for a new teacher evaluation system due to the fact that the current system provided little in the way of useful feedback, a high majority of teachers (over 95%) were given the highest ratings, and there was “a disconnect between current teachers’ evaluations and student achievement” (“Principal & Teacher Evaluation,” 2011). This was a collective effort between the regional offices of education, the Illinois State Board of Education, the Illinois Education Association, the Illinois Principals Association, the Illinois Board of Higher Education, area universities, the Illinois Association of School Administrators, and other professionals and professional organizations.

While the State of Illinois waited to see what kind of impact this new expanded teacher performance rating system has on the students, teachers, and school districts within its borders, it was beneficial to study what other states were doing across the nation to see if any information could be gleaned to inform efforts in Illinois. The National Comprehensive Center for Teacher Quality hosts interactive databases on its website which collect information on state teacher and principal evaluation policies (National Comprehensive Center for Teacher Quality [NCCTQ], 2012). This nearly complete national analysis provided evidence that many states across the nation are moving to performance evaluation ratings similar to that of Illinois.

According to the NCCTQ database information posted on their website on October 15, 2012 the following states mandated, recommended, or proposed a specific number of proficiency levels along with the following names for those levels:

Table 1

*State Proficiency Levels*

State	Proficiency Levels	Labels for Levels
Arizona	Recommends 4	Highly Effective Effective Partially Effective Ineffective
Arkansas	Recommends 4	Exemplary Proficient Basic Unsatisfactory
Colorado	Mandates 3 Recommends 4	Highly Effective Effective Partially Effective Ineffective
Connecticut	Mandates 4	Exemplary Proficient Developing Below Standard
Delaware	Mandates 4	Highly Effective Effective Needs Improvement Ineffective
District of Columbia	Mandates 4	Highly Effective Effective Minimally Effective Ineffective
Florida	Mandates 4	Highly Effective Effective Needs Improvement Unsatisfactory
Georgia	Recommends 4	Exemplary Proficient Developing/Needs Improvement Ineffective
Hawaii	Mandates 4	Highly Effective Effective Marginal Unsatisfactory
Idaho	Proposes 4	Distinguished Proficient Basic Not Proficient
Illinois	Mandates 4	Excellent Satisfactory

		Needs Improvement Unsatisfactory
Indiana	Mandates 4	Highly Effective Effective Improvement Necessary Ineffective
Iowa	Proposes 4	Highly Effective Effective Minimally Effective Ineffective
Kentucky	Proposes 4	Exemplary Accomplished Developing Ineffective
Louisiana	Mandates 2 Recommends 5	Highly Effective Accomplished Proficient/Effective Effective/Emerging Ineffective
Maine	Mandates 4	Not Specified
Maryland	Recommends 3	Highly Effective Effective Ineffective
Massachusetts	Mandates 4 (Teacher Practice)  Mandates 3 (Student Learning)	Exemplary Proficient Needs Improvement Unsatisfactory  High Medium Low
Minnesota	Recommends 4	Exemplary Proficient/Effective Developing Unsatisfactory/Ineffective
Missouri	Recommends 4	Distinguished Proficient Developing New/Emerging
New Jersey	Mandates 4	Highly Effective Effective Partially Effective Ineffective
New Mexico	Proposes 5	Exemplary Highly Effective Effective Partially Effective Ineffective

New York	Mandates 4	Highly Effective Developing Effective Ineffective
North Carolina	Recommends 5	Distinguished Accomplished Proficient Developing Not Demonstrated
Ohio	Mandates 4	Accomplished Proficient Developing Ineffective
Oklahoma	Mandates 5	Superior Highly Effective Effective Needs Improvement Ineffective
Pennsylvania	Proposes 4	Distinguished Proficient Needs Improvement/ Progressing Unsatisfactory
Rhode Island	Mandates 4	Highly Effective Effective Developing Ineffective
South Carolina	Recommends 4	Exemplary Proficient Needs Improvement Unsatisfactory
South Dakota	Mandates 4	Distinguished Proficient Basic Unsatisfactory
Tennessee	Mandates 5	Significantly Above Expectations Above Expectations At Expectations Below Expectations Significantly Below Expectations
Utah	Mandates 4	Highly Effective Effective TBD Ineffective
Virginia	Mandates 4	Exemplary Proficient Developing/Needs Improvement Unacceptable

Washington	Mandates 4	Distinguished Proficient Basic Unsatisfactory
West Virginia	Recommends 4	Distinguished Accomplished Emerging Unsatisfactory
Wisconsin	Recommends 3	Exemplary Effective Developing

---

*Note.* Adapted from National Comprehensive Center for Teacher Quality. (2012). *Databases on state teacher and principal evaluation policies (STEP database and SPEP database)*. Retrieved from <http://resource.tqsource.org/stateevaldb/>

Two states in particular that had been evaluating teachers in a model similar to that of the newly expanded Illinois system were Massachusetts and Florida.

Massachusetts and Florida both had been operating under an expanded four-tier model since at least the 2011-2012 school year. Therefore, many principals in these two states were able to compare and contrast from their old system to their expanded new system. These experiences could provide administrators in the State of Illinois the chance to learn from the experiences of the principals in Massachusetts and Florida in relation to their transition to a four-tiered teacher evaluation system.

The Massachusetts Board of Elementary and Secondary Education (MBESE, 2011) adopted the four-tier model on June 28, 2011. The overall summative ratings they assign to teachers for teaching practice are “exemplary,” “proficient,” “needs improvement,” and “unsatisfactory.” Prior to June 28, 2011, the Massachusetts State Code did not detail how many ratings to give and instead noted that the superintendent use “the regulations and principles adopted by the Board of Education and such consistent, supplemental performance standards as the school committee may require”

(MBESE, 1995). Massachusetts teachers also earn a second rating based on the educator's impact on student learning which is a rating of "low," "medium," or "high." This student learning rating was applied once the student learning measures have been identified and the necessary data have been available for two years (MBESE, 2011).

Florida first had their four-tier teacher evaluation model appear in their State Code in 2011. The ratings they assign to their teachers are "highly effective," "effective," "needs improvement" (or, for instructional personnel in the first three years of employment who need improvement, developing)," and "unsatisfactory" (Florida Legislative Statutes, 2011). Prior to this, Florida had no teacher evaluation rating distinctions in their State Code (Florida Legislative Statutes, 2010).

By conducting research on principals' perceptions of the teacher evaluation systems in these two states, the resulting data provided insight to inform Illinois regarding the perceived impacts of utilizing the expanded four-tier performance evaluation system that was recently adopted. By surveying all public school principals in Massachusetts and Florida from Kindergarten through 12<sup>th</sup> grade, an extensive body of data was collected to provide information to inform Illinois educational leaders based on the experiences of the principals within these states.

### **Research Questions**

The research questions attempted to focus on the various perceived impacts, intended and unintended, created by the state mandated expansion to a four-tier teacher performance evaluation rating model. In order to capture these perceived impacts the following research questions were researched and answered:

From the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e. “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e. “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

This study was considered a “natural experiment” as it described “a naturally-occurring contrast between a treatment and a comparison condition” (Shadish, Cook & Campbell, 2002, p. 17). There could be no manipulation of variables in this research and instead the study analyzed the manipulation in variables that naturally occurred as the

legislation in the states of Massachusetts and Florida mandated that principals change the rating system by which they evaluated teachers.

### **Significance of the Study to the Field of Educational Leadership**

The significance of this study to the field of educational leadership is that with Illinois shifting from a three-tier to a mandated four-tier performance evaluation system, it is important to understand how the expanded teacher evaluation ratings may impact teacher recognition, effectiveness, growth, remediation, and dismissal. As Illinois continues to expand its teacher performance rating categories a further analysis of the intended and unintended impacts of this expansion was important.

This dissertation studied the principals' perception of the impact of the expanding performance rating systems on teacher recognition, effectiveness, growth, remediation, and dismissal in Massachusetts and Florida. By understanding principal perceptions from Massachusetts and Florida in regards to measuring teacher performance and the resulting perceived impact on teacher recognition, effectiveness, growth, remediation, and dismissal, Illinois principals may learn important information as they continue to move forward.

### **Proposed Methodology**

The research methodology for the study was a cross sectional survey with quantitative and qualitative data. It was cross sectional because it was a snapshot in time of what people think or believe (Merriam, 2009). This resulted in a study bounded by the states of Massachusetts and Florida. The study surveyed all K-12 public school principals in the states of Massachusetts and Florida. The goal of the study was to understand the perceptions of principals from Massachusetts and Florida concerning the

intended and unintended consequences of expanding their respective teacher evaluation systems in order to inform Illinois principals about the perceptions of teacher rating usage and effectiveness on a number of levels between the models in Massachusetts and Florida and the new Illinois evaluation model.

In regards to the survey itself a number of questions were drafted and refined as the study neared implementation. The survey was piloted with 14 Loyola University School of Education administration and supervision doctoral students in the fall of 2012. The respondents took the survey and then provided feedback based on their experiences. This feedback included focused information around the validity of the survey in addition to effective formatting. The goal of any question included in the survey was to support the overarching research questions which were to explore and measure the perception of principals concerning expanded teacher performance ratings on teacher growth and effectiveness.

The vision of implementation for this study was to provide an electronic opportunity for principals to participate. The survey was generated using Survey Monkey® due to the security and confidentiality it provides. The survey was sent out via email to each K-12 public school principal in the states of Massachusetts (n= 1854) and Florida (n= 4533) that have active email addresses posted on their respective state department of education websites in March of 2013. The survey was accompanied by a letter making the participants aware of the study and the request to participate. This survey was sent out two more times via email over the course of a three week period with a follow-up request for participation.

The survey consisted of Likert scale and open-ended questions intended to collect information on the principals' perceptions of the impact of expanded teacher evaluation performance ratings on teacher recognition, effectiveness, growth, remediation, and dismissal. Demographic information was collected so as to further analyze these data upon survey completion. It was approximated that it would take participants fifteen minutes to complete the survey.

### **Areas of Related Literature**

In 2009 The New Teacher Project published *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. Its primary thesis posited why ineffective teachers in our schools go unaddressed. The report found that in districts that use a two rating scale, usually "satisfactory" or "unsatisfactory," that over 99% of the teachers received a "satisfactory" rating. In districts with more expansive rating scales 94% of teachers receive one of the top two ratings while less than 1% are rated "unsatisfactory." Despite these numbers 81% of administrators and 57% of teachers within this study reported there was a tenured teacher in their school who was performing poorly. Forty-three percent of teachers reported that there was a tenured teacher in their school that needed to be dismissed for poor performance. On the flip side, the report stated that excellence goes unrecognized as 59% of teachers and 63% of administrators say that their own district does not do enough to "identify, compensate, promote and retain the most effective teachers" (Weisberg, Sexton, Mulhern, & Keeling, 2009, p. 6).

The report stated that "Excellence goes unrecognized, development is neglected and poor performance goes unaddressed" (Weisberg et al., 2009, p. 10). The report

stated that the expectation was that if a school system employed a wider range of performance rating categories than the binary system that it would more accurately reflect teacher performance differences. What it found was that the districts studied rated the majority of teachers in the top category instead of just assigning this rating to the teachers who outperformed their peers. In these districts 70% of tenured teachers received the highest rating and the next 24% received the next highest rating (Weisberg et al., 2009).

In regards to recommendations, *The Widget Effect* stated that an effective performance evaluation system would have “multiple, district rating options that allow administrators to precisely describe and compare differences in instructional performance” (Weisberg et al., 2009, p. 27).

In June of 2010 a policy brief was released by the Consortium on Chicago School Research titled *Rethinking Teacher Evaluation: Findings from the First Year of the Excellence in Teaching Project in Chicago Public Schools*. The report cited a statistic that “83% of the state’s school districts had never rated a tenured teacher as “unsatisfactory”” (Sartain, Stoelinga & Krone, 2010). In Chicago Public Schools (CPS) in 2007-08 there were 91% of teachers who received a “superior” or “excellent” evaluation rating while at the same time 66% of CPS schools failed to meet state standards (Sartain et al., 2010).

In 2008-09 CPS began an evaluation pilot using the most recent Danielson *Framework for Teaching* (2007) in addition to their existing evaluation checklist. The pilot focused on Danielson’s two observable domains, The Classroom Environment and Instruction. The principals were able to rate teachers using the four Danielson levels of performance which were “Unsatisfactory,” “Basic,” “Proficient,” and “Distinguished”

(Sartain et al., 2010). In this pilot more teachers were identified as low-performing. In the sample, 8% of teachers received at least one “unsatisfactory” rating as opposed to 0.3% of teachers in CPS that had been rated “unsatisfactory” in past years (Sartain et al., 2010). The pilot program found that the Danielson *Framework for Teaching* had the potential for improving teacher evaluation systems and was “a reliable tool for identifying low-quality teaching” (p. 15).

In 2010 The New Teacher Project published a follow-up report to *The Widget Effect* which was titled *Teacher Evaluation 2.0*. This report proposed six design standards of a model teacher evaluation system. One of these design standards called for teacher evaluations to be comprised of multiple ratings. The report proposed that “each teacher should earn one of four or five summative ratings at the end of each school year: for example, “highly effective,” “effective,” “needs improvement” or “ineffective”” (Teacher Evaluation 2.0, 2010). The authors argued that this system can both give teachers a clear picture of their current performance and also be specific enough to allow for distinctions between levels and teacher differentiation across the district.

The report referenced the importance of having rating scales with no ambiguity. Rather, the performance evaluation ratings should have at least two levels at or above expectations and two levels below expectations. This type of rating system would provide clear information to teachers and assist administrators in making employment decisions (Teacher Evaluation 2.0, 2010).

### **Conceptual Framework**

Charlotte Danielson, one of the leading experts in the field of teacher evaluation, proposes rubrics that provide a framework for measuring effective teacher performance.

She stated that her rubrics for teacher performance “represent levels of performance of *teaching*, not of *teachers*” (Danielson, 2008). Danielson’s rubrics are based on four levels of performance which are “Unsatisfactory,” “Basic,” “Proficient,” and “Distinguished.” “Unsatisfactory” is characterized as performance that is resulting in harm being done to students, there is no learning taking place, and/or there is a chaotic environment. “Basic” is described as what one would expect from a new teacher, including inconsistent performance and inability to adjust lessons appropriately. “Proficient” and “Distinguished” teachers are characterized as experienced teachers who achieve high levels of student engagement and learning where students contribute to the success of the classroom. “Distinguished” in particular represents the highest level of meeting teacher standards which a beginning teacher would rarely attain (Danielson, 2008).

Danielson (2008) shared that in some school districts the evaluator is asked to assign ratings for each framework component in her rubric during an observation and she discourages this practice. One reason that she stated was that “performance is notoriously inconsistent, even among highly experienced teachers” (p. 52). Given that an observation is a microcosm of the entire year she argued that it should not be rated on its own. She instead argued that “it is unwise for an evaluation system to place high stakes on the outcome of any single observation of teaching” and “many other factors must be considered as well: informal observations of teaching, observations of other aspect of practice...and the consideration of artifacts that provide evidence of those aspects of teaching that cannot be observed at all” (p. 53).

Danielson (2008) also offered her professional opinion on how final evaluation ratings should be used. She discussed the different systems that exist including three rating systems (i.e. “Unsatisfactory,” “Satisfactory,” and “Outstanding”), other systems that are dichotomous with either a ‘does not meet’ or ‘meets or exceeds’ standard for example, and some systems that create algorithms to determine different final ratings. She said that although inadequate performance must be addressed these instances are rare. She said that beyond these low performers “as long as performance at least meets the district’s minimum standards it does not matter...to what extent the performance exceeds those standards” (p. 56). For those educators that meet or exceed expectations the focus should shift away from the rating and towards “identifying those aspects of practice that could be strengthened; that is, it shifts from summative to formative assessment of teaching” (p. 56). When school districts rate teachers on each component “then at least some teachers will put energy into challenging the rating, parsing the words, and arguing over evidence” (p. 56).

In 2011 Danielson released *The Framework for Teaching Evaluation Instrument*. This updated version added critical attributes for each level of performance for each component. It also provided possible examples to illustrate the meanings of each level on the rubric. Danielson made it clear that there were “absolutely no changes to the architecture of *The Framework for Teaching*” and therefore nothing would contradict those earlier versions (p. v).

The common argument against the dichotomous rating system is that it does not adequately appreciate the work of the teacher who has demonstrated high levels of performance and is only given a ‘meets expectations.’ Danielson (2011) agreed that the

teachers' work needs to be identified but argues that school districts should challenge themselves to find other ways to recognize "excellent" teaching rather than making it part of the evaluation system. The intent of this study was to inform educational leaders regarding the perceived impact of the expanded four-tier teaching evaluation rating system on teacher recognition, effectiveness, growth, remediation, and dismissal as perceived by principals in Florida and Massachusetts who now use this four-tiered model.

For each of the five areas being studied (recognition, effectiveness, growth, remediation, and dismissal) principal perceptions were analyzed using Danielson's (2008) four domains as a conceptual framework to understand these data. In surveying Massachusetts and Florida the four domains of: 1) planning and preparation, 2) classroom environment, 3) instruction, and 4) professional responsibilities were measured in relation to teacher recognition, effectiveness, growth, remediation, and dismissal using the four-tiered performance rating system.

### **Summary**

In summary, this study explored and measured the perceptions of Florida and Massachusetts principals concerning the perceived impact of teacher performance ratings on teacher recognition, effectiveness, growth, remediation, and dismissal using Charlotte Danielson's *Framework for Teaching* as a conceptual framework for the study. Following the adoption of a four-tier performance evaluation rating system in Illinois for 2012-13, this study analyzed principals' perceptions in the states of Massachusetts and Florida who adopted an expanded teacher rating system for teacher evaluation in the prior year in order to inform principals in Illinois.

## CHAPTER II

### LITERATURE REVIEW

#### **Introduction**

The goal of this literature review was to encapsulate current research to provide a context for this study which was to determine the perceived impact of expanded teacher performance ratings on the teacher evaluation process. In doing so the research intended to specifically answer the following research questions from the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?

- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

Through this literature review the following topics were researched and organized in order to capture major themes, studies, and topics related to the field of teacher evaluation in relation to the research questions noted above. This included a brief historical overview of teacher evaluation followed by a review of teacher evaluation through the lens of teacher recognition, teacher effectiveness, teacher growth, teacher remediation, and teacher dismissal. Next was an analysis of the relationship between principal and teacher in the evaluation process and the courageous and strategic conversations necessary to make teacher evaluation meaningful and effective. The perceived impact of the teacher evaluator on the fidelity of the teacher evaluation process was studied along with the perceived impact of evaluation on school culture and climate and how the evaluator can promote social justice through effective evaluation. Current trends and impacts on teacher evaluation were explored with specific emphasis on *Race to the Top* and Charlotte Danielson’s framework for effective teaching. Finally, teacher evaluation reform in Illinois, Massachusetts, and Florida was reviewed.

### **A Brief Historical Overview of Teacher Evaluation**

The concept of teacher evaluation is by no means a new one. In the 1709 document entitled *Reports of the Record of Commissions of the City of Boston* it was written that:

[It should] be therefore established a committee of inspectors to visit ye School from time to time, when as oft as they shall see fit, to Enform themselves of the methods used in teacher of ye Scholars and Inquire of their proficiency, and be present at the performance of some of their Exercises (Kyte, 1930, pp. 8-9).

Teacher supervision originated in the early 1700s and was largely done by clergy. The practice of clergy supervision extended through the mid-1800s (Marzano, Frontier, & Livingston, 2011).

In the mid-1800s more complex school systems started to develop in urban areas and “One teacher within a building was often selected to assume administrative duties. This ‘principal’ teacher ultimately grew into the role of building principal” (Marzano et al., 2011, p. 13). This was due to a growing demand for teachers who held content specific expertise and in turn a demand for administrators that could take on roles that were more specific and complex.

Approaching the end of the 19<sup>th</sup> century and into the beginning of the 20<sup>th</sup> century teacher evaluation entered a period of scientific management. This was characterized by the work of John Dewey and Frederick Taylor and their competing views. Dewey (1938, 1981) believed the function of schools was to promote citizenship and democratic ideals. Taylor (1911) took a factory production approach and believed that measuring specific behaviors would improve academic production. In using these two approaches to supervise and evaluate teachers Marzano et al. (2011) argued that “the two perspectives

were not described or perceived in a fashion that allowed for integration, and the tension between them continued through the Great Depression (p. 15).

Moving forward to 1921, the seminal text *Common Sense in School Supervision* was written by Charles A. Wagner. The idea of performance ratings for teachers emerged in his work. Wagner suggested the use of a scale such as the following five-tiered model:

- 5, seldom needs any suggestions from the supervisor; often supplies suggestions to the supervisor;
- 4, needs suggestions but always uses and adapts them wisely;
- 3, needs many suggestions, uses some, but seldom or never adapts them to her needs;
- 2, is helpless alone and must have suggestions about everything; seldom or never gets any suggestion used.
- 1, total failure; continuance impossible (Wagner, 1921, p. 149).

Wagner cited an example of criteria for effective evaluation in *The Fourteenth Yearbook of the National Society for the Study of Education* which was published in 1918. In this publication Professor Arthur Clifton Boyce cited his Methods for Measuring Teachers' Efficiency as 45 items under five major categories:

- Personal Equipment – 14 items
- Social and Professional Equipment – 12 items
- School Management – 4 items
- Technique of Teaching – 10 items
- Other – 5 items (p. 146)

Following World War II there was a shift away from the scientific approach. As a reaction away from the industrial man, the teacher evaluation process focused more on the teacher as an individual instead of a cog in the educational machine. Elsie Coleman (1945) captured this new sentiment by stating that “the first fundamental in understanding the teacher is...that the teacher is a person, different from every other person” (p. 165).

This stage led to an elaboration on what effective teaching represented and also gave rise to the importance of the classroom observation. Matthew Whitehead (1952) highlighted the new emphasis of classroom observation by stating “administrators should pay more attention to the chief aim of education – effective teaching” (p. 106) and also stressed that administrators should follow up “the visitation with a conference, and in having the principal see the importance of remaining the entire period” (p. 102).

Such philosophies that came out of the post-WWII era set the stage for the clinical supervision era of teacher evaluation that would follow in the 1960s and 1970s and still remains in different forms today. Morris Cogan and his colleagues in the Harvard Master’s of Arts in Teaching program were credited with creating the clinical supervision model which paralleled a model often seen in teaching hospitals (Cogan, 1973; Marzano et al., 2011). This model was refined by Robert Goldhammer (1969) who developed the five-phase process which included the pre-observation conference, classroom observation, analysis, supervision conference, and analysis of the analysis.

In the 1980s Madeline Hunter moved to the forefront of supervision and evaluation. Among her contributions was the Hunter model of lesson design (1980, 1984) which emphasized that an effective lesson should consist of an anticipatory set, objective and purpose, input, modeling, checking for understanding, guided practice, and independent practice. As Marzano (2011) stated “if clinical supervision was the prescribed structure of supervision, Hunter’s seven-step model...became the content of the preconference, observation, and postconference” (p. 20).

Moving to present day, the methods for measuring teacher efficiency introduced by Boyce in the early 1900s mirrors two widely used teacher performance models. The

first is Charlotte Danielson's *Framework for Teaching* (2013) and the second is Robert Marzano's *Knowledge Base for Teaching* (2011).

Danielson's model has four domains which total 22 components and 86 elements.

They are built around the following domain categories:

- Planning and Preparation – 6 components, 23 elements
- The Classroom Environment – 5 components, 15 elements
- Instruction – 5 components, 18 elements
- Professional Responsibilities – 6 components, 20 elements (Danielson, 2013).

Marzano's model also has four domains which comprise 60 specific elements.

Those domains are categorized by the following titles:

- Classroom Strategies – 41 elements
- Planning and Preparing – 8 elements
- Reflecting on Teaching – 5 elements
- Collegiality and Professionalism – 6 elements (Marzano et al., 2011).

What this brief historical overview captured was though teacher evaluation has evolved over time there is also much that has remained the same. The profession continues to explore the most effective criteria and methods for evaluating teachers. This includes a recent emphasis on factoring in student learning and growth as a significant component of teacher evaluation. Teacher performance is increasingly being measured by student growth and attainment of learning standards in a variety of ways from state to state. If history is any indicator of future practices the field of education will continue to evolve and refine this process without significant deviation.

### **Teacher Evaluation and Teacher Recognition**

In Wagner's previously referenced 1921 work he stated that "Besides putting the good teacher on her mettle, this scale permits every teacher to feel that her individuality has a real chance to demonstrate itself and to secure recognition" (p. 149). He also says

that “Even the teacher who is rated “5” or “excellent,” may be satisfied by knowing that the supervisor regards her work excellent” (p. 147).

The recognition of teacher performance through the evaluation process is a concept with multiple perspectives and opinions. Secretary of Education Arne Duncan stated in a 2010 speech that “our system of teacher evaluation...frustrates teachers who feel that their good work goes unrecognized and ignores other teachers who would benefit from additional support” (The New Teacher Project, 2010, p. 1). This sentiment was echoed in the *Widget Effect* which argued that “When all teachers are rated good to great, those who are truly exceptional cannot be formally identified” (Weisberg et al., 2009, p. 6).

The value teachers and administrators place on teacher evaluation performance ratings may play an important role in how effective the ratings and evaluations are in recognizing teacher performance. On one hand, unions that reflect an industrial orientation “are likely to insist on having only two ratings – satisfactory and unsatisfactory – in order to discourage more nuanced judgments by administrators” (Johnson & Donaldson, 2006, p. 132). This, however, raises the argument presented in the *Widget Effect* that “In districts that use binary evaluation ratings...more than 99 percent of teachers receive the satisfactory rating” (Weisburg et al., 2009, p. 6). In such a system ratings do little to identify or reward excellence as the “average effort becomes the bar for the mark of excellence” (Weisburg et al., 2009, p. 13).

When school districts extend beyond the binary system of teacher performance ratings Weisburg et al. (2009) argue there is still minimum impact on recognizing excellence as “94 percent of teacher receive one of the top two ratings and less than 1

percent are rated unsatisfactory” (p. 6). Compound this with the fact that “only 42 percent of teachers agree that evaluation allows accurate assessment of performance” and you have a teacher evaluation system that could be criticized as coming up short with recognizing teachers through performance ratings (p. 14).

The National Council on Teacher Quality argued that teacher evaluation systems ‘teach to the middle’ much like schools tend to do with students. In regards to teacher recognition for excellent performance they stated “with evaluation tools neither designed nor implemented with an eye towards identifying the most talented educators or those who struggle. The reality is that there is huge variation in teacher performance...But the disregard for performance in education has bred massive dysfunction...” (National Council on Teacher Quality, 2011, p. 3).

Thomas J. Sergiovanni (1992) stressed the importance of recognition for performance under the framework of teacher motivation. He connected teacher recognition and motivation by presenting three rules of motivation in leadership. His first rule applied most directly to teacher recognition which was ‘what gets rewarded gets done.’ He goes on to argue that this was only extrinsic motivation though and to tap into intrinsic motivation or moral motivation the rules are that ‘what is rewarding gets done’ and ‘what is good gets done’ (p. 25). Although ‘what gets rewarded gets done’ may only lead to extrinsic motivation it still supports the notion that recognizing teachers for their efforts and performance will lead to further motivation. Sergiovanni does point out that the opposite was also true in that what does not get rewarded does not get done. He claimed that in an extrinsically motivated model “workers become increasingly

dependent on rewards and on their leaders to motivate them. ‘What gets rewarded gets done’ discourages people from becoming self-managed and self-motivated” (p. 25).

Many arguments have been raised over time that merit or performance pay will provide the necessary recognition to further motivate teachers. This research has been met with mixed outcomes. Marzano (2011) claimed that teachers are not motivated by money, “However, they *are* motivated by recognition of expertise. This idea is not new. National Board certification is designed singularly for this purpose” (p. 9). This argument of the value of teacher recognition was furthered by Danielson (2007) who argued that her own framework for teaching “offers educators a means of communicating about excellence” (p. 6). Danielson’s ideas will be explored in greater depth later in this chapter.

### **Teacher Evaluation and Teacher Effectiveness**

Rewinding and revisiting Wagner (1921) once again, he stated “It is also unmistakably true, that teachers marked on [a rating scale] basis are surer to anticipate their marks and to feel they have been fairly treated than if the mark be an estimate of excellence described by a word with no indication why that quality assigned” (p. 149).

The impact that teacher evaluations, and the potential ratings that accompany the evaluation, have on teacher effectiveness is another philosophical argument to be debated. The question revolves around how accurately the teacher evaluation system measures teacher effectiveness and what potential outcomes on the future effectiveness of the teacher can be derived from the evaluation process.

The overarching premise of the *Widget Effect* “describes the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher. This

decades-old fallacy fosters an environment in which teachers cease to be understood as individual professionals, but rather as interchangeable parts” (Weisburg et al., 2009, p. 4). If this is true then the teacher evaluation process is not recognizing and promoting teacher effectiveness.

In the *Teacher Evaluation 2.0* report (2010) it offered a solution to recognizing and promoting teacher effectiveness. This report proposed that teachers should earn a summative rating on a four or five category scale. The argument was that “this number of categories is large enough to give teachers a clear picture of their current performance, but small enough to allow for clear, consistent distinctions between each level” (National Council on Teacher Quality, 2011, p. 7).

Sergiovanni (2007) took a step back from the importance of the teacher evaluation ratings and argued that a more differentiated approach to teacher evaluation was necessary to recognize and promote teacher effectiveness. He broke these teacher evaluation purposes down to either quality control or professional improvement and depending upon the category the focus should be different.

When evaluating for quality control Sergiovanni (2007) claimed the following elements should be present so that teachers may measure up to standards, criteria, expectations and procedures:

- the process should be formal and documented;
- the criteria should be explicit;
- the standards should be uniform for all teachers;
- the criteria should be legally defensible as being central to basic teaching competence;
- the emphasis should be on teachers meeting requirement of minimum acceptability;
- the responsibility for evaluation should be in the hands of the evaluators.

When evaluating for professional improvement the following elements should be present so that teachers may increase their understanding and enhance their teaching practices, standards, criteria, expectations, and procedures in a different form:

- the process should be informal;
- the criteria should be tailored to the needs and capabilities of individual teachers;
- the criteria should be considered appropriate and useful to teachers before they are included in the evaluation;
- the emphasis should be on helping teachers reach agreed-upon professional development goals;
- the teachers should assume major responsibility for the process by engaging in self-evaluation and peer evaluation, and by obtaining evaluation information from students (p. 235).

Capturing the essence of what makes a teacher effective can be an elusive process which can make the process of recognizing this effectiveness through the evaluation rating process equally elusive. As Parker Palmer (2007) stated, “good teaching cannot be reduced to technique; good teaching comes from the identity and integrity of the teacher” (p. 10). He goes on to explain that the ability to connect with students, be vulnerable, infuse their personal identity into their work, and focus on the service of learning are the qualities of effective teaching.

According to Michael Fullan (2010), the focus on teacher effectiveness is not about the ratings but instead is about the incentives that promote effectiveness in teaching. He argued that these incentives “are related to working conditions that enable groups to accomplish impressive results that have high moral values” (p. 88). That list was comprised of the following components:

- Good salaries;
- Decent surroundings;
- Positive climate;

- Strong induction;
- Extensive professional learning;
- Opportunity to work and learn from others (job embedded and otherwise);
- Supportive, and even assertive, leadership about the agenda;
- Getting helpful feedback;
- Reasonable class size;
- Long-term collective agreements (4 years);
- Realizable moral purpose.

Fullan said that tapping into these components when linked with a moral purpose are what motivated teachers in “helping them achieve dramatic success with students that they did not think could learn” (p. 89).

### **Teacher Evaluation and Teacher Growth**

Framing the teacher evaluation process and teacher growth around Wagner’s (1921) work he stated the following regarding the teacher evaluation rating: “Hence the mark must convey a double significance. It must show degree of shortcoming, or need of improvement. To be helpful and corrective it must also show the means to be used to effect improvement. The second showing is just as necessary and important as the first” (p. 148).

The debate around the impact of the teacher evaluation process on teacher growth is a rich one. Revisiting the *Widget Effect* (2009), the report stated that there was a failure to identify specific teacher developmental needs due to a lack in ability to assess variations in teacher instructional effectiveness. The report stated, “In fact, 73 percent of teachers surveyed said their most recent evaluation did not identify any development areas” (p. 6). Although this report called for more accurate teacher evaluation ratings to further identify teacher effectiveness it acknowledged that, “In theory, even if virtually all teachers are rated as good or great, their evaluations could provide them with valuable

feedback they could use to improve their instructional practice. However, that theoretical potential currently goes unrealized and teachers are too often denied both the knowledge and the opportunity to improve” (p. 14). The concerns around the impact of teacher evaluation promoting teacher growth are compounded by their findings that “Only 43 percent of teachers agree that evaluation helps teachers improve” (p. 14).

One of the challenges around promoting teacher growth is that research related to what is considered best practice is not without argument. Sergiovanni (2007) stated that “There is no conclusive and incontrovertible research that any specific teacher behavior or any set of teacher behaviors causes learning to take place in any specific student” (p. 298). He asserted that evidence pointed to weak correlations between some behaviors and increased scores on basic competency tests. Other experts in the field refute this position and some of these stances are captured later in this chapter but the reality is that researchers constantly grapple with what constitutes best practice for teachers which makes providing feedback through evaluation a challenge (Danielson, 2013; Marzano et al., 2011; Schmoker, 2011; Wiggins & McTighe, 1998).

Given what we do know to be best instructional practice the focus on teacher growth should turn to professional development. According to Zepeda (2007) expenses around professional development should be significant as “school districts across the United States spend over 80 percent of their budgets on staff salaries. Given this expenditure, opportunities for professional development need to be elevated as a top priority if schools are to realize maximum return on this investment” (p. 28).

Zepeda (2007) argued that professional development and teacher evaluation need be linked and even embedded throughout the supervision and evaluation cycle and “there

are ways to bring together supervision, professional development, evaluation, and other activities such as peer coaching and mentoring. The real charge for principals is to unify these efforts” (p. 26). Furthermore, she stated that professional planning needs to be a calculated process by principals and they “are in an advantageous position to identify professional development needs and to provide follow-up support teachers need to implement new skills into their daily practices” (p. 37).

Sergiovanni (2007) reinforced the importance of professional development to promote teacher growth. He presented the 80/20 quality rule as a balance schools should reach in relation to teacher evaluation. As he defined the 80/20 rule he stated “When more than 20 percent of supervisory time and money is expended in evaluation for quality control or less than 80 percent of supervisory time and money is spent in professional improvement, quality schooling suffers” (p. 236). He argued that teacher evaluation needed to follow this framework in order to promote quality schooling.

Stronge, Gareis, and Little (2006) present another argument to support the importance of linking teacher evaluation and teacher growth. They stated that the “primary goal of a teacher evaluation system should be to encourage continuous growth and improvement at an individualized level by collecting and analyzing pertinent data and utilizing those data as the foundation for meaningful feedback” (p. 38). Although this argument tied in the rationalization for a flexible compensation model to further promote this growth they still provided further support for using the evaluation process as an ongoing method of continuous improvement and growth.

Israel and Kersten (2007) further endorsed the concept of linking teacher evaluation and teacher growth by sharing their belief that “to create systematic impact,

the educational leader must understand, translate, and then bundle multiple theories of best educational practice into a practical model that fits his/her school and/or school district” (p. 45). This supported the previous work of Kersten and Israel (2005) which argued that supervision and evaluation of teachers alone was not the answer to promoting whole school improvement from the perception of educational leaders. Instead, their 2007 study concluded that collaboration, instead of mandates, played a key role in effective teacher growth through staff development and “no substantial educational improvement effort is possible without the active support and positive involvement of the faculty and administration” (p. 55).

The *Teacher Evaluation 2.0* report (2010) stated that “Evaluations should provide all teachers with regular feedback that helps them grow as professionals; no matter how long they have been in the classroom” (p. 1). The report acknowledged that there needs to be consequences for poor performance but this should not be the primary function of evaluations and instead “Good evaluations...encourage a school culture that prizes excellence and continual growth” (p. 2).

### **Teacher Evaluation and Teacher Remediation**

Returning to Wagner (1921), he stated “if [the rating] shows that the teacher has neglected to use suggestions given by the supervisor, it necessarily indicates that improvement of the mark can be earned by a more sympathetic and intelligent use of suggestions given” (p. 148). Teacher remediation is a necessary component to the teacher evaluation process when suggestions for growth are not enough, recurrent patterns emerge in poor performance, and/or the need for improvement becomes more urgent. There are certain obligations, but also direct benefits, in initiating a remediation

process with a teacher as opposed to seeking termination. Jackson (2008) asserted that “Yes, it is important to eliminate mediocre or poor teaching, but the best way to get rid of mediocre or poor teaching is to help those teachers improve” and this was done by focusing on a culture of growth and improvement with appropriate support to improve teacher practice (p. 9).

The process of remediation takes significant work on the part of the evaluator and the teacher. Revisiting a previously cited quote from Secretary of Education Arne Duncan from his 2010 speech he stated, “our system of teacher evaluation...frustrates teachers who feel that their good work goes unrecognized and ignores other teachers who would benefit from additional support” (The New Teacher Project, 2010, p. 1). According to the *Widget Effect* (2009), “Even when performance is clearly an issue...evaluators fail to invest significant time monitoring instruction” (p. 21). In order for an evaluation system to be effective the remediation component of the teacher evaluation process cannot be ignored.

According to the *State of States* report (2011) many states are redesigning their remediation practices associated with their teacher evaluation systems. These systems are setting clear expectations and processes for teachers who receive poor evaluation ratings. The report stated that “The most promising policies on this front spell out both the kinds of interventions required and a specific time period within which ineffective teachers should have an opportunity to demonstrate improvement or be dismissed” (p. 23).

Nolan and Hoover (2005) provided specific guidance on what an effective remediation process entailed. When a marginal teacher is identified for evaluation the

stakes are raised and “because the competence of a veteran teacher is in question, evaluation takes center stage” and “although the goal of the process is improvement of performance, the procedures used must also comply with legal requirements for teacher dismissal if that option proves necessary” (p. 296). The process of remediation is truly a balancing act as the ultimate goal for the evaluator is to promote teacher growth but he/she also must proceed in such a manner so that dismissal may be an option if the remediation process proves to be unsuccessful. If the message was not already clear to the teacher, the remediation process becomes that clear communication that improvement is needed in their performance.

As Nolan and Hoover (2005) argued, “Remediation as a goal makes sense from a variety of viewpoints – ethical, organizational, legal, and economic...the primary reason for investing time and effort to remediate the performance of marginal teachers is thus moral and ethical. It is simply the right thing to do” (p. 300). The argument is that successful remediation prevents the damages that are associated with dismissal while promoting growth in pursuit of teacher effectiveness. Successful remediation also sends important institutional messages that “First, poor teaching performance is not acceptable. Second, the district is prepared to help teachers improve their performance and will work hard at doing so” (p. 300).

### **Teacher Evaluation and Teacher Dismissal**

Wagner’s (1921) teacher rating system acknowledged that at some point remediation is no longer an option and dismissal is the only viable solution. He defined this by a rating of 1 which was accompanied by the description of total failure where continuance is impossible.

In 2009 President Barack Obama commented that “If a teacher is given a chance or two chances or three chances but still does not improve, there is no excuse for that person to continue teaching. I reject a system that rewards failure and protects a person from its consequences. The stakes are too high. We can afford nothing but the best when it comes to our children’s teachers and the schools where they teach” (Weisburg et al., 2009, p. 2). Unfortunately, teacher dismissal today is a tenuous process that is wrought with philosophical, moral, and legal stances and implications.

Teacher dismissal through the teacher evaluation process could be viewed as a viable option to address President Obama’s stance but it often is not that easy or that widely used. The Illinois Small Newspaper Group found in a 2005 report that “83% of [Illinois] school districts had never rated a tenured teacher as “unsatisfactory.” School systems as diverse as Denver, Chicago, Atlanta, and San Francisco rarely dismiss low-performing teachers – often less than 1 percent of teachers in any given year” (Sartain et al., 2010, p. 1). A March 6, 2010 edition of Newsweek cited the following information related to teacher dismissal:

In most states, after two or three years, teachers are given lifetime tenure. It is almost impossible to fire them. In New York City in 2008, three out of 30,000 tenured teachers were dismissed for cause. The statistics are just as eye-popping in other cities. The percentage of teachers dismissed for poor performance in Chicago between 2005 and 2008 (the most recent figures available) was 0.1 percent. In Akron, Ohio, zero percent. In Toledo, 0.01 percent. In Denver, zero percent. In no other socially significant profession are the workers so insulated from accountability. The responsibility does not just fall on the unions. Many principals don't even try to weed out the poor performers (or they transfer them to other schools in what's been dubbed the “dance of the lemons”). Year after year, about 99 percent of all teachers in the United States are rated “satisfactory” by their school systems; firing a teacher invites a costly court battle with the local union (Thomas & Wingert, 2010).

Johnson and Donaldson (2006) took a closer look at the topic of teacher dismissal through the lens of the collective bargaining agreement and they found that “state laws generally set a higher standard for dismissing tenured than nontenured teachers because the courts have determined that tenured teachers have a vested property right to a job under the 14<sup>th</sup> amendment. Thus, districts must provide due process for all tenured teachers who are dismissed” (p. 131).

Johnson and Donaldson (2006) further pointed out that a school district’s success in dismissing poor teachers, or improving the performance of others, is limited by laws, contracts, and unions. The variability of these constraints dictates the level of success teacher evaluators encounter in pursuing teacher dismissal as follows:

- 1) School officials may find the negotiated procedures for reviewing teachers’ performance either reasonable or burdensome.
- 2) Contracts may include a rating scheme that provides detailed feedback for all teachers about their performance or distinguishes only among the competent and incompetent.
- 3) Union officials may decide to aggressively defend all members who receive negative evaluations, or they may do no more than protect the procedural rights of their members, as the collective bargaining laws require.

Where a district lands in regards to these variables “can create a situation in which principals regularly assess all teachers and move to dismiss those who are ineffective, or one in which teacher dismissal is a contentious, politically charged event that principals rarely undertake” (p. 131).

The reality is that the benefits are at times outweighed by the costs of pursuing a teacher dismissal (Kersten & Israel, 2005; Stronge, Gareis, & Little, 2006; Zepeda, 2007). Some teacher evaluators avoid the process of dismissal altogether because “they view the dismissal process as overly time consuming and cumbersome, and the outcomes for those who do invest the time in the process is uncertain” (Weisburg et al., 2009, p. 17). Compound this with the financial and emotional strains it places on the evaluator and the process is far from appealing. This leads some administrators to be “reluctant to move for dismissal because they have heard horror stories of dismissal attempts that were extremely costly and were eventually overturned by the courts because of some minor procedural error” (Nolan & Hoover, 2005, p. 316).

Teacher dismissal is not an impossible task despite the odds that may seem to be against its success. Nolan and Hoover (2005) argued that a district that has clear evaluation standards that are consistently applied and documented in cooperation with providing “the teacher with notice of the deficiencies and significant attempts at improvement, and has accorded the teacher the appropriate due process procedures, the chances of losing a dismissal case are quite slim” (p. 316).

Perhaps most importantly is that the pursuit of teacher dismissal is doing what is right for students, regardless of the hurdles and barricades an evaluator encounters along the way. This will be explored in greater depth later in this chapter. As Sergiovanni (2007) argued, “The outcome of evaluation for quality control should be the protection of students and the public from incompetent teaching. Unquestionably this is an important outcome and a highly significant responsibility for principals and other supervisors, as well as teachers” (p. 235).

### **The Relationship between Principal and Teacher**

The relational dynamic between principal and teacher is one that requires a delicate balance in the teacher evaluation process. It is not a far stretch to argue that the reason so many teachers are rated in the category of “satisfactory” to “excellent” is because their evaluators fear implications to their relationship with teachers when they rate them otherwise. In Chicago’s *Excellence in Teaching Project*, which started in 2008-2009, the principals surveyed were able to identify unsatisfactory teaching practices, “however, when using the high end of the scale, principals inflated their ratings...Principals acknowledged this tendency, pointing to the need to preserve relationships with teachers who had previously received the highest possible evaluation rating” (Sartain et al., 2010, p. 7). Although not always an easy task, Sergiovanni (2007) argued that supervision and evaluation needed to remain separate entities in order to uphold trust and collegiality. He stated that separation of these responsibilities make sense as “Evaluating teachers can dampen, if not betray, the collegiality and trust that are needed for teacher learning to take place” (p. 168).

Revisiting Wagner (1921) one last time, he stated that: “if supervision is to lead to a teacher rating that shall win and hold the respect of teachers, however, it must eliminate some of the present crudities and contradictions, like our arbitrary values and variety of opinions” (p. 154). In order to address those crudities and contradictions teacher evaluators need to make evaluation an ongoing process with frequent conversations and “if teachers are surprised by their summative evaluation rating, something is wrong with the evaluation process” (The New Teacher Project, 2010, p. 8). Instead, these regular conversations should capture observations of classroom performance, professional

growth and development goals, support to meet those goals and student progress. When this relationship is modeled correctly “Teachers and instructional managers should come away from these conversations with a shared understanding of what the teacher needs to focus on” (The New Teacher Project, 2010, p. 8).

Further promoting the idea of the impact of a positive and productive relationship was Jackson (2008) who argued that “In the same way that students work best with teachers with whom they have a positive relationship, teachers work best with leaders with whom they have a positive relationship” (p. 10). She argued that this is the heart of strategic conversations which will be explored further in the following section. Jackson said that strategic conversations “help you establish trust and maintain it – even when you are sharing really difficult feedback. When teachers feel safe, they are more likely to take the steps they need to improve” (p. 10).

The current political landscape has further complicated the teacher and principal dynamic around teacher relationships and teacher evaluation. As evaluation starts to carry greater clout in regards to teacher pay, teacher retention, benefits, as well as the obstacles around linking evaluation to student performance, the relationship becomes convoluted. Conley and Glasman (2008) asserted that the dimensions of public, political, bureaucratic, and market accountabilities “has placed teacher evaluation as one of the pivotal controversial foci of the debate involving both accountability-related policies and accountability-related student outcome- based measurement and evaluation” (p. 68). As more implications become tied to the evaluation process, “The result is that teachers may fear that evaluation is less about personal improvement involving professional growth and more of a political hurdle” (p. 68) and “this development exacerbates teachers’ fears

of being evaluated for responsibilities and activities that they control only minimally” (p. 81). Stronge et al. (2006) asserted that support from a critical mass of key constituent groups including policymakers, administrators, teacher, and community members was important to implementing an effective evaluation and compensation system but that ultimately, “compensation is a decision of taxpayers and policymakers” (p. 160).

Conley and Glasman (2008) argued that in order to control the fear it is first important to identify its sources. Their study narrowed the sources down to three primary areas:

1. the prospect of losing control and autonomy in one’s work
2. working in an atmosphere of organizational rigidity and inflexibility
3. failing to be continued in one’s profession and/or lacking a sense of continuous skill development or career progress

In response to these fears they propose the following solutions for teachers:

- 1) union participation
- 2) altering the adversarial tone of evaluation
- 3) furthering collaboration and teamwork
- 4) joint principal and teacher analysis of student learning

They emphasized that due to the multiple sources of fear “a single “Band-Aid” remedy that would redress fear appears unlikely” but through the solutions proposed above they may “remove a bit of fear, giving more certainty and enhancing teacher evaluation within a school” (p. 75). Furthermore, the importance of teachers and evaluators working together to overcome obstacles must mean that they are together, “codesigners of work

environments that provide sufficient resources to meet the increased demands of teaching. Indeed, fear could lead to bland and cautious teaching outcomes” (p. 81).

Sergiovanni (1992) took a different approach to addressing an effective teacher and principal relationship in regards to teacher evaluation. He framed the relationship between principal and teacher through his model of sources of authority (see Figure 2). He argued that there are five different sources of authority that could guide leadership policy and practice which are: bureaucratic authority, psychological authority, technical-rational authority, professional authority, and moral authority.

Sergiovanni (1992) argued that professional and moral authorities are the “sources of authority on which to base leadership practice. Neither one is management- or leadership-intensive, and both create a response in teachers that come from within, rather than being imposed” (p. 31). He stated that there is a place for psychological, bureaucratic, and technical-rational authority in leadership “but that its place should be to provide support for professional and moral authority. The latter two [professional and moral] should be the primary bases for leadership practice” (p. 33).

When moral authority can be attained then the leader can foster a school grounded in ethics as the foundation for effective decision making. Robert Starratt (2012) furthered the idea of moral leadership through his beliefs on cultivating an ethical school. Starratt wrote that “just as medical ethics is concerned with promoting the good of its professional practice, which is physical health; just as business ethics is supposed to be concerned with promoting the public and individual good involved in trade, commerce, and contracts, just so one would expect educational ethics to be grounded in the particular good involved in teaching and learning” (p. 108). Starratt described the ethics that

teachers should be promoting to include student self-understanding in relation to the world and becoming a productive member of this world. He framed the role of ethical teaching “to focus very intentionally on the *proactive* pursuit, cultivation, and support of those goods of learning in and for a democratic community and polity” (p. 108).

Table 2

*Sergiovanni's Sources of Authority*

Source of Authority	Characteristics
Bureaucratic	Leader/Teacher relationship is a clear top-down model. Goals of teacher do not align with those of supervisor. Predetermined standards. Comply or face consequences.
Psychological	Leader/Teacher relationship is more give and take in defining goals. School climate is more congenial. Teachers extrinsically motivated by rewards to reach compliance.
Technical-Rational	Leader/Teacher relationship relies on logic and science. Facts and evidence supersede values and beliefs. Teachers comply due to what they believe to be best practice and scientifically rational.
Professional	Leader/Teacher relationship driven by respect for professional norms. Recognizes multiple approaches can reach desired outcome. Craft knowledge and personal expertise lie at the foundation of decision making. Teacher autonomy and shared values hold each other accountable for performance.
Moral	Leader/Teacher relationship based on shared commitments and interdependence with shared norms and values. Community is created that is driven by what is right and good. Leader does not dictate but instead norms govern behavior of teachers.

*Note.* Adapted from Sergiovanni, T. (1992). *Moral leadership: Getting to the heart of school improvement*. San Francisco, CA: Jossey-Bass, Inc.

## **Courageous and Strategic Conversations**

At the core of effective principal and teacher relationships is the ability to have courageous and strategic conversations around teacher evaluation. As Jackson (2008) stated “Because strategic conversations are often uncomfortable, you [the evaluator] and the person with whom you are conversing may naturally want to avoid them” (p. 12). A foundation of strategic conversations is truth as “it’s hard to share honest feedback with your colleagues when that feedback is not positive...Unless you can provide honest feedback to teachers, they cannot act on your feedback and improve their practice” (p. 12).

Scott (2004) asserted that “If your stomach flips at the thought of confronting someone’s behavior, you’re in excellent company” (p. 136). He argued that this apprehension is rooted in past experiences with confrontation that include the following fears being realized:

- A confrontation could escalate the problem rather than resolve it.
- I could be rejected.
- I could lose the relationship.
- Confronting the behavior could force an outcome for which I am not prepared.
- I could incur retaliation.
- The cure could be worse than the disease.
- I could be met with irrationality or emotional outbursts.
- I might hurt his or her feelings.
- I could discover that I am part of the problem.

But he further argued that not addressing these fears and confronting the conversation head on could lead to:

- The problem could escalate rather than be resolved.
- I could be rejected.
- I could lose the relationship.
- Emotions could escalate until someone blows up.

By not have strategic and courageous conversations around teacher evaluation the evaluator is likely only prolonging the inevitable. By doing so Scott believed the undesirable behavior and need for a conversation “will just take longer, and the results will likely occur at the worst possible moment, when we are least expecting it, with a huge price tag attached” (p. 136).

An effective strategy in which to frame these courageous conversations is based on the foundation that the evaluator does not have to be doing all of the heavy lifting. Jackson stressed that “Strategic conversations emphasize problem solving among staff. The instructional leader is not the problem solver; the instructional leader facilitates problem solving among teachers” (Jackson, 2008, p. 10). In this model teachers have responsibility for their professional growth and the evaluator helps to facilitate this growth.

Scott elaborated further on this concept and stresses the importance of clearly describing the behavior that needs to be confronted so that the other person understands the concern and can also explain his or her point of view. When this type of conversation is fostered, “learning is provoked, and most people are willing to take action once they have gained a new understanding. Such conversations enrich relationships” (Scott, 2004, p. 139).

Conversations to foster learning are not always be met with great success and in such cases a different type of conversation may be necessary. Nolan and Hoover (2005) suggested a more direct approach which progresses from “gentle persuasion to improve, to increase negative feedback, to threats of an unsatisfactory evaluation, to an actual unsatisfactory rating, followed by counseling that it is time to exit the profession before

the administrator has no choice but to move for dismissal” (p. 315). With such an approach the end goal remains the same to have honest, courageous, and strategic conversations around teacher evaluation with the best interests of students at the core of these conversations.

### **Teacher Evaluator Impact**

The impact of the teacher evaluator in the process of evaluation cannot be underscored. The competency of this individual plays a critical role in a fair teacher evaluation process for the sake of teachers and the students they serve. Unfortunately, studies of evaluation effectiveness by teacher evaluators have found “unrepresentative sampling, biased reporting, disruptions caused by the classroom visit, and limitations on the principal imposed by misleading or truncated reporting systems such as checklists and narrow anecdotal category systems” (Peterson, 2004, p. 61).

Peterson’s (2004) study captured recommendations for high-quality teacher evaluation based upon strong principal leadership. These recommendations included the following for principals:

- can help teachers to actively participate in teacher evaluation;
- can encourage teachers to become knowledgeable about the need for good teacher evaluation and defensible data gathering;
- can effectively advocate in their district for development of sources of data and observation checklists;
- can help individual teachers to experiment with data sources;
- can help teachers guide credible collection of evidence of quality for their own evaluations;
- can help to educate teachers in the need for including student achievement data;
- can encourage staff development to help teachers select and write effective assessments;
- can advocate for teacher-dominated panels to advocate for, and monitor, teacher evaluation in their school district. (p. 72)

In a 2005 study, Kersten and Israel found that teacher evaluators believed they could make a difference in teaching and learning despite obstacles standing in the way. They stated that “Teacher evaluation, when conducted appropriately, has the potential to improve teaching and learning” but in the pursuit of a more comprehensive evaluation system “that depicts the true nature of teaching and learning, we may have created a monster” (p. 62). The primary impediment cited in the study was how time intensive the teacher evaluation process has become and how this in turn provides fewer opportunities for leaders to collaborate with teachers to improve classroom instruction. The authors claimed that if teacher evaluators have the opportunity to provide “increased communication opportunities, data-driven targeted staff development, peer coaching and mentoring, as well as principal demonstration teaching, they can improve instruction in the classroom” (p. 62).

In addition, the teacher unions expect effective teacher evaluation because they too recognize the impact of the evaluator on the teacher. Johnson and Donaldson (2006) stated that teacher unions expect principals to be fair and responsible in their teacher evaluations and in return, “privately, union leaders often explain that they have agreed not to defend members they know to be weak unless these teachers’ procedural rights are violated” (p. 133). Unfortunately, Kersten and Israel’s (2005) research data showed that “principals perceive unions as not trusting the more complex, subjective teacher evaluation methods that are currently considered best practice” (p. 62). This reality could interfere with teacher evaluation process as it could impact the relationship, and in turn the impact, of the teacher evaluator’s perceived effectiveness.

A 2009 study by Kimball and Milanowski showed that significant variation could be found in the validity of teacher evaluator ratings. They argued that “this suggests that estimates of criterion-related validity should be interpreted with caution and that the quality of ratings may vary considerably across evaluators” (p. 67). In response to this finding they suggested extensive evaluator training to pursue greater rating validity. The consequences of the negative impact of teacher evaluator ratings are significant and significant variation could lead to a scenario where “teachers could receive consequences that are not justified” (p. 35).

When ratings do accurately align with teacher performance the impact of the teacher evaluation should correlate with the teacher’s impact on student achievement. Specifically, if there is a relationship between the “teacher behaviors specified by the system and student learning, an accurate set of ratings will exhibit a stronger relationship with student achievement than an inaccurate set” (Kimball & Milanowski, 2009, p. 39). The authors argued that with the increased attention on teacher accountability; the skill and decision-making ability of the teacher evaluators in regards to teacher performance has been raised to a critical level.

The National Council on Teacher Quality State of the States report (NCTQ, 2011) claimed that “It is clear that performance-based evaluations will require more from evaluators and observers of teacher performance than they have in the past” (p. 31). In response to this claim the report asserted that states now need to make significant investments in evaluator training or poor implementation will cripple the evaluation system.

## **Culture and Climate**

Past and present teacher evaluation practices can have a significant impact on the culture and climate of school communities at a local, state, and national level in a positive or negative way. Present day examples can be found in Indiana, Michigan, and Florida where the states are required to notify parents if their child is placed in an ineffective teacher's classroom. Rhode Island on the other hand has set the goal of ensuring by 2015 that no student be taught for more than one year by an ineffective teacher, although it does not publicly notify parents who those teachers are that are ineffective. The National Council on Teacher Quality (2011) argued against the first practice stating that "If a district has evidence that a teacher is ineffective, state policy should provide the means for the district to take the necessary steps to remove the individual from the classroom, not humiliate the teacher" (p. 35).

Given the current climate around student performance being tied to teacher evaluation the match between teacher and student takes on even another dynamic. It is not only the parents that are worried about where the students are being placed but also, "One of the things causing teachers considerable trepidation is the concern on how they will be matched with students" (NCTQ, 2010, p. 36). Under this construct of student performance measurements impacting teacher evaluation teachers now want to be matched with students who have the greatest potential for growth and achievement. Although opinions may differ regarding what that student profile may look like teachers understand that they could be penalized for being paired with students who may be the most difficult to teach in regards to promoting growth and achievement.

Charlotte Danielson (2007) has weighed in on the climate surrounding teacher evaluation and stated that “An environment of high-stakes accountability only exacerbates teachers’ levels of stress... Teachers are under enormous external pressure, as never before, to prepare their students for productive lives in the knowledge economy and success in externally mandated assessments” (p. 5). Furthermore, she asserted that “It is well known that fear shuts people down” and an evaluation system in which teachers do not feel threatened is the system in which they will learn the most (p. 182).

Sergiovanni (2007) supported Danielson’s position of creating a collaborate teacher evaluation process to address climate and culture by stating “Much of the discomfort concerning evaluation can be eliminated, however, if it is treated as a community exercise in self-governance, as a way for the school community to maintain and strengthen its commitment to learning” (p. 297). Although Sergiovanni believed that some of the conflict around teacher evaluation is healthy, providing a collaborative approach would help to reduce the tension associated with the process.

Toch and Rothman (2008) also agreed that a comprehensive evaluation system was the key to promoting a positive and productive school climate. They stressed the importance of scoring rubrics, multiple classroom observation by multiple evaluators, and the consideration of student work and teacher reflections. The authors argued that this combination can “contribute much more to the improvement of teaching than today’s drive-by evaluations or test scores alone. And they contribute to a much more professional atmosphere in schools” (p. 13).

## **Promoting Social Justice**

Finding direction and clarity in light of what is the complicated landscape that currently surrounds the teacher evaluation process may be found through the context of making decisions based on the concept of promoting social justice. If the compass for what is right is based on promoting student achievement and growth for all students then the focus for evaluators needs to be on effectively evaluating teachers due to the link that teacher expertise has on student achievement.

Marzano et al. (2011) shared that the focal point of teacher supervision and evaluation should be to promote growth around teachers' pedagogical skills with the end goal of these skills promoting student achievement. He declared that "One incontestable fact in the research on schooling is that student achievement in classes with highly skilled teachers is better than student achievement in classes with less skilled teachers" (p. 2). Figure 3 illustrates the point that as teacher skill increases the predicted student growth is impacted to a much greater level. Given these projections Marzano et al. asserted that the implication for teacher supervision and evaluation was clear in that "its primary purpose should be the enhancement of teacher expertise" (p. 2).

Schmoker (2011) argued the same point from the opposite direction stressing that "In education...the general underperformance of schools can be directly attributed to a failure to implement three simple, well-known elements: a common curriculum, sound lessons, and authentic literacy" (p. 9). He continued his argument by claiming we have not done enough as an educational system to promote the impact of these three elements and says that even if they were just implemented 'reasonably well' they would have a significant impact on student achievement.

Table 3

*Teacher Expertise and Student Achievement*

Teacher Skill Percentile Rank	Predicted Percentile Gain for Student at the 50 <sup>th</sup> Percentile	Predicted Percentile Rank for Student
50 <sup>th</sup> percentile	0%	50 <sup>th</sup> percentile
70 <sup>th</sup> percentile	8 %	58 <sup>th</sup> percentile
90 <sup>th</sup> percentile	18%	68 <sup>th</sup> percentile
98 <sup>th</sup> percentile	27%	77 <sup>th</sup> percentile

*Note.* Adapted from Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria: ASCD.

The call to improve teacher evaluation systems for the betterment of students was also argued by Toch and Rothman in their 2008 report entitled *Rush to Judgment*. They argued that factors such as a lack of accountability for performance, staffing practices that reduce the significance of the teacher evaluation, union ambivalence, and the emphasis of teacher credentials over teacher performance “have resulted in teacher evaluation systems throughout public education that are superficial, capricious, and often don’t even directly address the quality of instruction, much less measure students’ learning” (p. 1).

These arguments for the importance of quality teaching are further argued by Marshall and Oliva (2010) who directly tied the importance of incorporating social justice into the equation. They shared that there is clear evidence that not all schools, or even all students within a school, have equal access to quality teachers. When these quality teachers are not equally distributed “Students of color and students from low-

income homes most often have less experienced teachers, teachers with less education and training, and more teachers teaching without certification and/or outside their areas of expertise” (p. 266).

Sergiovanni (1992) further promoted the importance of social justice in relation to promoting student growth by sharing that “the virtuous school believes that every student can learn, and it does everything in its power to see that every student does learn” (p. 112). He stressed that this means providing learning conditions that do not impede learning and addressing these problems instead of just accepting them. He further supported this argument by stating that “every parent, teacher, student, and administrator is viewed as an interdependent member of the school as covenantal community and that every action taken in the school must seek to advance the welfare of this community” (p. 106). In doing so, all of the members of the school community need to be treated with equality, dignity, and fair play to the benefit of the school community as a whole.

Starratt (2012) shared that a virtuous school is one that cultivated ethical character in such a way that it is not an ‘add-on’ but instead “it should permeate the purpose and process of every element in the school...all elements and aspects of the school life should be managed with and should reflect an ethic of care, justice, and critique” (p. 141). Starratt shares that in an ideal situation the ethical school would be cultivated by district administrators, building administrators, teachers, counselors, and other professional staff.

### **Current Trends and Impacts on Teacher Evaluation**

As has previously been highlighted there are significant changes going on across the nation in regards to teacher evaluation trends and impacts. Some of these are driven

by initiatives such as the *Race to the Top* while others are impacted by models for best teaching practice such as Charlotte Danielson's *Framework for Teaching*.

The 2011 *State of States* report captured the following recent national trends in regards to teacher evaluation:

- Thirty-two states and D.C. made changes to their state teacher evaluation policies in the past three years.
- In the past 2 years the number of states requiring annual evaluations of all teachers went from 15 to 24.
- In the past 2 years the number of states tying student achievement data to teacher evaluations grew from 15 to 23 (NCTQ, 2011, p. 21).

Marzano et al. (2011) further observed that “since the turn of the 21<sup>st</sup> century, emphasis has shifted from supervision to evaluation, as well as from teacher behavior to student achievement” (p. 25). Tucker and Stronge (2005) pushed for the importance of student growth measures being part of the teacher evaluation process and as an important source of feedback on educator effectiveness by stating “given the clear and undeniable link that exists between teacher effectiveness and student learning, we support the use of student achievement information in teacher assessment” (p. 102).

Peterson (2004) also touted the importance of linking achievement data with teacher evaluation and stated that this was the desire of legislatures, parents, and taxpayer groups who are pushing for greater school quality. Peterson shared that “educators who expect support for public education have a burden to make their case with the best objective evidence about student learning. A teacher evaluation system that does not strive for pupil gain does not get the best data that are available” (p. 64).

The concept of expanded teacher performance ratings have also been a frequently explored topic as of recent. The *State of States* report asserted that, “the only clear right answer at present on the number of performance levels seems to be more than two” (NCTQ, 2011, p. 21). The report stated that four or five levels may be the better options. Having four options forces raters to use discretion and having five options provides greater opportunity for differentiation in the ratings.

### **Race to the Top**

*Race to the Top* was a product of the American Recovery and Reinvestment Act (ARRA) of 2009 which was signed into law on February 17, 2009, by President Obama. The goal of ARRA was to stimulate the economy, support job creation, and invest in critical sectors such as education. ARRA provided the *Race to the Top* fund with \$4.35 billion in competitive grants that rewarded states for education innovation and reform. States would be rewarded for reform that led to significant student outcome improvements, improvements in high school graduation rates, and better preparing students for college and career preparation (U.S. Department of Education, 2009, p. 2).

Four specific areas of core education reform were specifically targeted for states to ambitiously implement in their plans:

- 1) adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy;
- 2) building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction;
- 3) recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and
- 4) turning around our lowest achieving schools (U.S. Department of Education, 2009, p. 2).

The federal *Race to the Top* (RTT) competition “spurred unprecedented action among the states to secure a share of \$4 billion” (NCTQ, 2011, p. 5). *Race to the Top* funding opportunities prompted states across the nation to reform and realign their teacher evaluation systems so as to be eligible for this funding source. Naturally, this funding came with many strings attached. One criterion under the grant requirements, according to the *Race to the Top* Executive Summary (2009) was, “improving teacher and principal effectiveness based on performance” (p. 3). Specifically, points were awarded to applicants who “differentiate effectiveness using multiple rating categories that take into account data on student growth...as a significant factor” (U.S. Department of Education, 2009, p. 9). The application did not elaborate on specifically how many performance rating categories were to be used or what descriptors should be used to identify the ratings.

Several of the early RTT winners put plans in place to address the performance criteria. Delaware, Florida, Massachusetts, Rhode Island, Tennessee, and the D.C. Public Schools all instituted plans that required annual evaluations of teachers and included student learning evidence, “not as an option, but as the *preponderant criterion* for assessing teacher effectiveness” (NCTQ, 2011, p. 5). This study examined the principal perceptions of two of these early winners, Florida and Massachusetts.

### **Charlotte Danielson’s Framework for Teaching**

One model that has emerged nationwide to support effective teaching practices and to serve as a template for teacher evaluation is Charlotte Danielson’s *Framework for Teaching*. Marzano et al. (2011) acknowledged Danielson’s model by sharing that

“Given its past and current popularity, the Danielson model must be the reference point for any new proposals regarding supervision and evaluation” (p. 23).

Danielson arrived at her framework for teaching through an analysis of empirical studies and theoretical research related to improved student learning. The framework is comprised of 4 domains, 22 components, and 76 descriptive elements (see Figure 1).

Danielson (2007) argued that the origin of identifying elements of professional practice was rooted in the work of Madeline Hunter (1982) and the research around process-product and cognitive science. She claimed that Hunter “was one of the first educators to argue persuasively that teaching is not only an art but also a science; some instructional practices are demonstrably more effective than others” (p. 7). Process-product research reinforced this message by finding relationships between some teaching practices and their impact on student achievement (Dewey, 1933; Gardner, 1983; Hunter, 1982).

Danielson (2013) explained that the idea of a framework for professional practice was not unique to education and “other professions – medicine, accounting, and architecture, among many others – have well-established definitions of expertise and procedures to certify novice and advanced practitioners. Such procedures are the public’s guarantee that the members of a profession hold themselves and their colleagues to high standards of practice” (p. 2). The existence of this framework allows for use by educators and the community as a whole. The framework was touted by Danielson to be used for a variety of professional practices including: the preparation of new teachers, the recruitment and hiring of teachers, a road map for novices, guidance for experienced

professionals, a structure for focusing improvement efforts, and communication with the larger community.

<p><b>Domain 1: Planning and Preparation</b></p> <p>1a. Demonstrating knowledge of content and pedagogy</p> <ul style="list-style-type: none"> <li>▪ knowledge of content and the structure of the discipline</li> <li>▪ knowledge of prerequisite relationships</li> <li>▪ knowledge of content-related pedagogy</li> </ul> <p>1b. Demonstrating knowledge of students</p> <ul style="list-style-type: none"> <li>▪ knowledge of child and adolescent development</li> <li>▪ knowledge of the learning process</li> <li>▪ <i>knowledge of students' skills, knowledge and language proficiency</i></li> <li>▪ <i>knowledge of students' interests and cultural heritage</i></li> <li>▪ <i>knowledge of students' special needs</i></li> </ul> <p>1c. Setting instructional outcomes</p> <ul style="list-style-type: none"> <li>▪ value, sequence and alignment</li> <li>▪ clarity</li> <li>▪ balance</li> <li>▪ suitability for diverse learners</li> </ul> <p>1d. Demonstrating knowledge of resources</p> <ul style="list-style-type: none"> <li>▪ resources for classroom use</li> <li>▪ resources to extend content knowledge and pedagogy</li> <li>▪ resources for students</li> </ul> <p>1e. Designing coherent instruction</p> <ul style="list-style-type: none"> <li>▪ learning activities</li> <li>▪ instructional materials and resources</li> <li>▪ instructional groups</li> <li>▪ lesson and unit structure</li> </ul> <p>1f. Designing student assessments</p> <ul style="list-style-type: none"> <li>▪ congruence with instructional outcomes</li> <li>▪ criteria and standards</li> <li>▪ design of formative assessments</li> </ul>	<p><b>Domain 2: Classroom Environment</b></p> <p>2a. Creating an environment of respect and rapport</p> <ul style="list-style-type: none"> <li>▪ teacher interaction with students</li> <li>▪ student interactions with one another</li> </ul> <p>2b. Establishing a culture for learning</p> <ul style="list-style-type: none"> <li>▪ importance of the content</li> <li>▪ expectations for learning and achievement</li> <li>▪ student pride in work</li> </ul> <p>2c. Managing classroom procedures</p> <ul style="list-style-type: none"> <li>▪ management of instructional groups</li> <li>▪ management of transitions</li> <li>▪ management of materials and supplies</li> <li>▪ performance of non-instructional duties</li> <li>▪ supervision of volunteers and paraprofessionals</li> </ul> <p>2d. Managing student behavior</p> <ul style="list-style-type: none"> <li>▪ expectations</li> <li>▪ monitoring of student behavior</li> <li>▪ responses to student misbehavior</li> </ul> <p>2e. Organizing physical space</p> <ul style="list-style-type: none"> <li>▪ safety and accessibility</li> <li>▪ arrangement of furniture and use of physical resources</li> </ul>
<p><b>Domain 4: Professional Responsibilities</b></p> <p>4a. Reflection on Teaching</p> <ul style="list-style-type: none"> <li>▪ accuracy</li> <li>▪ use in future teaching</li> </ul> <p>4b. Maintaining accurate records</p> <ul style="list-style-type: none"> <li>▪ student completion of assignments</li> <li>▪ student progress in learning</li> <li>▪ non-instructional records</li> </ul> <p>4c. Communicating with families</p> <ul style="list-style-type: none"> <li>▪ information about the instructional program</li> <li>▪ information about individual students</li> <li>▪ engagement of families in the instructional program</li> </ul> <p>4d. Participating in a professional community</p> <ul style="list-style-type: none"> <li>▪ relationships with colleagues</li> <li>▪ involvement in a culture of professional inquiry</li> <li>▪ service to school</li> <li>▪ participation in school and district projects</li> </ul> <p>4e. Growing and developing professionally</p> <ul style="list-style-type: none"> <li>▪ enhancement of content knowledge and pedagogical skill</li> <li>▪ receptivity to feedback from colleagues</li> <li>▪ service to profession</li> </ul> <p>4f. Showing professionalism</p> <ul style="list-style-type: none"> <li>▪ integrity and ethical conduct</li> <li>▪ service to students</li> <li>▪ advocacy</li> <li>▪ decision making</li> </ul>	<p><b>Domain 3: Instruction</b></p> <p>3a. Communicating with students</p> <ul style="list-style-type: none"> <li>▪ expectations for learning</li> <li>▪ directions and procedures</li> <li>▪ explanations of content</li> <li>▪ use of oral and written language</li> </ul> <p>3b. Using questioning and discussion techniques</p> <ul style="list-style-type: none"> <li>▪ quality of questions</li> <li>▪ discussion techniques</li> <li>▪ student participation</li> </ul> <p>3c. Engaging students in learning</p> <ul style="list-style-type: none"> <li>▪ activities and assignments</li> <li>▪ grouping of students</li> <li>▪ instructional materials and resources</li> <li>▪ structure and pacing</li> </ul> <p>3d. Using assessment in instruction</p> <ul style="list-style-type: none"> <li>▪ assessment criteria</li> <li>▪ monitoring of student learning</li> <li>▪ feedback to students</li> <li>▪ student self-assessment and monitoring of progress</li> </ul> <p>3e. Demonstrating flexibility and responsiveness</p> <ul style="list-style-type: none"> <li>▪ lesson adjustment</li> <li>▪ response to students</li> <li>▪ persistence</li> </ul>

*Note.* Adapted from Danielson, C. (2013). *The framework for teaching evaluation instrument*. Princeton, NJ: The Danielson Group.

*Figure 1.* Charlotte Danielson's Domains and Elements of the Framework for Teaching

The four domains of: Planning and Preparation, The Classroom Environment, Instruction, and Professional Responsibilities are intentionally balanced with five or six components for each domain. Danielson (2007) stated that “the different domains and components represent areas of roughly equal ‘size’ or heft.” One domain is not noticeably larger than the others, nor does one component within a domain reflect a much larger part of a teacher’s responsibility within that domain than do the other components” (p. 23).

There are levels of performance that are attached to the different domains. These levels are “unsatisfactory,” “basic,” “proficient,” and “distinguished” (see Figure 5). Danielson (2007) made a point to note that “It is important to recognize that the levels are levels of performance of *teaching*, not of *teachers*” (p. 39). She suggested that the levels of performance be applied to mentoring, coaching, and professional growth.

Danielson (2007) rarely referenced evaluation in her framework for teaching descriptions but asserted that “if the framework is to be used for supervision and evaluation, it is essential that it describe actual practice. That is, it is possible, from the manner in which the statements are written, to imagine ways in which a teacher might demonstrate skill in that area” (p. 23).

In reality, the Danielson model is widely used across the nation for teacher evaluation and in 2011 and 2013 she released two new editions of the framework that were titled *The Framework for Teaching Evaluation Instrument*. The 2011 version offered updated rubric language, the addition of critical attributes to accompany each level of performance for each component, and possible examples for each level of performance for each component. These additions were in response to the framework

being one of the models for the *Measures of Effective Teaching* (MET) research project led by Bill and Melinda Gates. The 2013 version was released with the Common Core State Standards (CCSS) in mind. The enhancements to the framework in this version were largely focused on Common Core ‘possible examples’ section of the framework instead of changes in regards to the rubric language or critical attributes.

Table 4

*Danielson Levels of Performance Descriptors*

Level of Performance	Descriptors
Unsatisfactory	<ul style="list-style-type: none"> <li>• no understanding of concepts underlying the component</li> <li>• growth and development by working on fundamental practices</li> <li>• below the standard of “do no harm”</li> <li>• time for supervisor intervention</li> </ul>
Basic	<ul style="list-style-type: none"> <li>• understands concepts underlying component and attempts to implement</li> <li>• implementation is sporadic, intermittent, or otherwise not entirely successful</li> <li>• additional reading, discussion, visiting classrooms of other teachers, and experience will lead to becoming proficient</li> <li>• considered minimally competent</li> <li>• level of most newer teachers</li> <li>• no harm is being done to students</li> <li>• enhancement of skill is important</li> </ul>
Proficient	<ul style="list-style-type: none"> <li>• clearly understands concepts underlying component and implements well</li> <li>• level of most experienced and capable teachers</li> <li>• thoroughly know content, students, curriculum, and possess broad repertoire of strategies and activities</li> <li>• sophisticated understanding of classroom dynamics</li> <li>• mastered the work of teaching while working to improve their practice</li> </ul>
Distinguished	<ul style="list-style-type: none"> <li>• master teachers that make a contribution to the field, both in and outside their school</li> <li>• classrooms consist of a community of learners, highly motivated students, engaged and assuming considerable responsibility for their own learning</li> <li>• classroom seems to be running itself</li> <li>• a place to visit, but don’t expect to live there</li> <li>• goal of all teachers</li> </ul>

*Note.* Adapted from Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*, 2<sup>nd</sup> ed. Alexandria, VA: ASCD.

## **Teacher Evaluation Reform in Illinois**

Illinois applied three times for *Race to the Top* funds and was denied the first two times with the third application being accepted. During Phase One, which was submitted in January of 2010, Illinois cited their commitment to robust teacher evaluations by the enactment of the Performance Evaluation Reform Act (PERA) of 2010. The State of Illinois *Race to the Top* Application for Initial Funding (2010) stated at the beginning of the section on performance evaluation systems that, “Teacher and principal evaluation in Illinois is broken.” It goes on to state that in three of the state’s largest districts they found that 92.6% of teachers were rated “superior” or “excellent,” 7% were rated as “satisfactory,” and 0.4% were rated “unsatisfactory.”

The grant proposal explained the expansion from three rating categories for teachers to four rating categories with the addition of a “Needs Improvement” category which was added to the categories of “Excellent,” “Proficient,” and “Unsatisfactory” beginning in 2012-13. PERA also eliminated the prior ability of school districts to obtain waivers to bypass this rating system which over 60 schools had obtained in the past to often implement a binary rating system (“State of Illinois Initial Application,” 2010).

The State of Illinois *Race to the Top* Phase Two application was submitted in June of 2010 after PERA was approved. In addition to reiterating the information and proposed actions steps from Phase One the application also presented specific detail regarding the role of the Performance Evaluation Advisory Council (PEAC) in implementing the PERA initiatives. One of the tasks assigned to PEAC was to begin the process of defining state standards of evaluation feedback that were both timely and constructive (“State of Illinois Phase Two,” 2010).

Illinois submitted Phase Three of the *Race to the Top* in December of 2011. Little had changed in regards to teacher performance ratings in this latest version (“State of Illinois Phase Three,” 2011). On December 22, 2011, Illinois was officially notified that it received funding from *Race to the Top* Phase Three. Funds were awarded in the amount of \$42,818,707 (“Race to the Top: Phase 3 Award Letter, Illinois, 2011). Thirty-five school districts in Illinois agreed to be part of this grant, including the Chicago Public School district which was awarded just over \$19,000,000 of the total grant award (“Illinois Race to the Top Phase 3: Allocations for Participating LEAs,” 2012).

The Performance Evaluation Reform Act was approved by Governor Pat Quinn on January 15, 2010. Beyond the Act’s function of expanding the teacher performance rating categories, it required student growth to be a “significant” factor in teacher and principal performance categories. This growth function was implemented for all principals statewide in 2012-2013 and by all teachers in Illinois school districts by an incremental process which will be completed by 2016-2017 (Performance Evaluation Reform Act [PERA], 2010).

Section 5 of the Performance Evaluation Reform Act began with a focus on the State’s findings. These included the concept that effective teachers and effective school leaders play a major role in student achievement. The findings then go on to focus on how the Illinois school district performance evaluation systems “fail to adequately distinguish between effective and ineffective teachers and principals” (PERA, 2010). The findings continued by citing that in a recent study of the three largest Illinois school districts that “out of 41,174 teacher evaluations performed over a 5-year period, 92.6% of

teachers were rated “superior” or “excellent,” 7% were rated “satisfactory,” and only 0.4% were rated “unsatisfactory” (PERA, 2010).

In Illinois, school districts can choose to adopt the state-designed classroom observation model which is adapted from Charlotte Danielson’s *Framework for Professional Practice*, develop their own model, or create a hybrid of the two. District administrators and teacher union representatives must work together to develop the evaluation system for their district. The Danielson framework is the default model if no collective decision can be agreed upon (Illinois State Board of Education, 2013).

Along with a revision to the performance ratings used in Illinois, PERA also stated that the performance evaluation system must go beyond measuring professional competencies and must also assess student growth. In doing so it is under the direction of PERA that the State of Illinois and individual school districts “must ensure that performance evaluation systems are valid and reliable and contribute to the development of staff and improved student achievement outcomes” (PERA, 2010). In order to establish student growth measures as a component of teacher performance ratings each district was directed to create a joint committee composed of an equal representation selected by the district and its teachers.

In regards to performance ratings it was stated in Section 24A-5 that “each teacher in contractual continued service is evaluated at least once in the course of every two school years and probationary teachers must be evaluated annually. However, any tenured teacher...whose performance is rated as either “needs improvement” or “unsatisfactory” must be evaluated at least once in the school year following the receipt of such rating” (PERA, 2010). Those that are given a rating of “needs improvement” will

begin the process of a professional development plan with a focus on the areas that need to improve and district supports in these identified areas. Those receiving an “unsatisfactory” will begin the process of a remediation plan in the event that the deficiencies are deemed ‘remediable’ (PERA, 2010). If a teacher with a “needs improvement” or “unsatisfactory” fails to complete any part of their remediation plan with a rating of “satisfactory” or better they are to be dismissed in accordance with Section 24-12 or 34-85 of the School Code.

Senate Bill 7 Public Act 97-0008, also known as Ed Reform, elaborated further on how performance evaluation categories impact teachers, school districts, and the State. Section 24-1.5 further emphasized the new importance of the performance ratings. Relevant experience will not be considered as a factor in filling a vacant teaching position unless other factors (certifications, qualifications, merit and ability – including performance evaluations) are equal.

Senate Bill 7 also incorporated the significance of the performance evaluation ratings into the process for teachers to obtain tenure. These probationary teachers can obtain tenure in one of three ways. The first path to tenure is if the teacher obtains a rating of at least “proficient” in their fourth school term in addition to at least a “proficient” in the second or third school term. A teacher that fails to meet these requirements is mandated by law to be dismissed at the end of the fourth school term. The second path to tenure is if the teacher earns three consecutive terms of “excellent.” The third possible path to tenure is if the teacher had previously achieved tenure in a different district and received at least a “proficient” for his/her two most recent post-PERA evaluations followed by a rating of “excellent” for the first two school terms in

their new district. Probationary teachers may still be non-renewed or dismissed by school boards with certain provisions (Illinois Pension Code Senate Bill 7, 2010).

Reductions-in-force (RIFs) are another major component of Senate Bill 7 that incorporate performance evaluation categories into the decision making process. All teachers now are placed into one of four groups for every position they are qualified to teach based on performance evaluation categories. When RIFs do occur the dismissals will begin with Group 1 and move toward Group 4 sequentially from there. To simplify a more elaborate process Group 1 consists of probationary teachers who have not yet been evaluated while Group 2 consists of teachers with either a “needs improvement” or “unsatisfactory” performance rating on either of their last two ratings. Group 3 consists of teachers who received a “proficient” rating on both of their last two evaluations and Group 4 consists of teacher who either received “excellent” on their last two performance evaluations or received “excellent” ratings on two of their last three evaluations and a third rating of “proficient” during that span. Length of teacher service only plays a role as a tiebreaker within the different groups and does not supersede the groups (Illinois Pension Code Senate Bill 7, 2010).

Section 21-23(a) of Senate Bill 7 addressed suspension or revocation of certificates and has specific language regarding unsatisfactory ratings. It defined incompetency as “two or more school terms of service for which the certificate holder has received an “unsatisfactory” rating on a performance evaluation....within a period of 7 school terms of service” (Illinois Pension Code Senate Bill 7, 2010). At this point the decision on whether or not to take action against a teacher’s certificate lies within the purview of the State Superintendent (see Table 5).

Table 5

*Illinois Poor Teacher Evaluation Consequences*

Policy for Assisting Teachers Who Receive Poor Evaluations	Teachers are Eligible for Dismissal Based on Poor Evaluations	Policy for Dismissing Ineffective Teachers
When a teacher receives a rating of needs improvement they must be placed on a professional development plan to address those areas. Those rated unsatisfactory must be placed on a remediation plan.	Yes.	Classroom ineffectiveness is specifically identified as grounds for dismissal. For teachers placed on remediation plans for poor performance that receive a subsequent unsatisfactory performance rating within three years, the school district may forego remediation and seek dismissal.

*Note.* Adapted from National Council on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Retrieved from the National Council on Teacher Quality website: [www.nctq.org/p/publications/docs/nctq\\_stateOfTheStates.pdf](http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf)

Two states in particular that have been evaluating teachers in a model similar to that of Illinois are Massachusetts and Florida. Massachusetts and Florida both have been operating under an expanded four-tier model since at least the 2011-2012 school year. The teacher evaluation models in all three states include default models that incorporate evaluation rubrics from Charlotte Danielson, Robert Marzano, and/or Kim Marshall but allow for the discretion of individual districts to adopt the model of their choosing with certain parameters in place to guide these decisions. Therefore, many principals in these two states were able to compare and contrast from their old system to their new system given their additional experience with the new model in comparison to Illinois. This

provided administrators in the state of Illinois the chance to learn from the experiences of the principals in Massachusetts and Florida.

The research questions of this study focused on the various perceived impacts, intended and unintended, created by state mandated expansion to a four-tier teacher performance evaluation rating models. In order to capture these perceived impacts the following research questions were researched and answered:

From the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?

- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

### **Teacher Evaluation Reform in Massachusetts**

In August of 2010 Massachusetts was awarded \$250 million through *Race to the Top*. In the grant Massachusetts highlighted four main objectives for its grant implementation. The four areas were great teachers and leaders, curricular and instructional resources, concentrated support in low-performing schools, and college and career readiness. The state focused part of its efforts on great teachers and leaders by concentrating on improved teacher and principal effectiveness based on performance and in turn created the Massachusetts Task Force on the Evaluation of Teachers and Administrators (U.S. Department of Education, 2012).

The Massachusetts Board of Elementary and Secondary Education (MBESE, 2011) adopted the four-tier model on June 28, 2011. The overall summative ratings that evaluators assign to teachers for teaching practice are “exemplary,” “proficient,” “needs improvement,” and “unsatisfactory” (see Table 6). Prior to this the Massachusetts State Code did not detail how many ratings to give and instead noted that the superintendent use “the regulations and principles adopted by the Board of Education and such consistent, supplemental performance standards as the school committee may require” (MBESE, 1995).

Table 6

*Massachusetts Teacher Rating Definitions*

Rating	Description
Exemplary	Practice is consistently and significantly above proficiency on the Standard or overall
Proficient	Practice demonstrates skilled performance on the Standard or overall
Needs Improvement	Practice demonstrates lack of proficiency on the Standard or overall
Unsatisfactory	Practice demonstrates lack of competence on the Standard or overall

*Note.* Adapted from Massachusetts Department of Elementary and Secondary Education (2011). *Building a breakthrough framework for educator evaluation in the commonwealth*. Retrieved from <http://www.doe.mass.edu/eval/breakthroughframework.pdf>

Massachusetts teachers also earn a second rating based on the educator's impact on student learning which is a rating of "low," "medium," or "high" (see Figure 2). This student learning rating will be applied once the student learning measures have been identified and the necessary data has been available for two years (MBESE, 2011).

Summative Rating	Exemplary	1-Year Self-Directed Growth Plan	2-Year Self-Directed Growth Plan		
	Proficient				
	Needs Improvement		Directed Growth Plan		
	Unsatisfactory		Improvement Plan		
		Low	Moderate	High	
		Rating of Impact on Student Learning			

*Note.* Adapted from Massachusetts Department of Elementary and Secondary Education (2011). *Overview of the new Massachusetts educator evaluation framework*. Retrieved from <http://www.doe.mass.edu/eval/101511Overview.pdf>

Figure 2. Massachusetts Teacher Rating Chart

The ratings on teacher practice uses observations, artifacts, and other evidence which are relevant to the standards in determining the final rating. Each educator must include at least one professional growth goal and one student learning goal in their evaluation plan. The student learning goal can include statewide, district, and classroom-based measures (MBESE, 2011).

Beginning in the fall of 2013 there was a second rating added (low, moderate, or high) based on the impact on student learning. This is a completely separate rating and is measured “based on a review of trends and patterns using at least two measures that are comparable at the state or district level across grades and subjects” (MDESE, 2011, p. 7).

The state has a system of growth plans that align with their summative rating process. By combining the summative rating and the impact on student learning the educator in Massachusetts arrives upon a specific plan. This plan may either be “Developing,” “Self-Directed Growth,” “Directed Growth,” or “Improvement.”

The Developing Educator Plan is designated for teachers without their professional teacher status or an educator new to an assignment if the evaluator so chooses. These plans are typically one year or less and are developed by the educator and evaluator. The Self-Directed Growth Plan is for experienced educators rated “proficient” or “exemplary” on either a one or two-year plan depending upon their rating for student learning. The Directed Growth Plan is for educators who have been designated as needing improvement and is developed by the educator and evaluator. Lastly, the Improvement Plan can last between 30 calendar days and one year and is developed by the evaluator for educators who are rated “unsatisfactory” (MDESE, 2011, p. 13).

The Standards of Professional Practice for Teachers were adapted from the National Board for Professional Teaching Standards (NBPTS) and the Interstate New Teachers Assessment and Support Consortium Standards (INTASC). There are five standards with indicators which align with the teacher evaluation rubrics of Charlotte Danielson and Kim Marshall. Districts in Massachusetts have the option to adopt these indicators or prescribe to others as long as they “describe essential activities related to the standards” (Massachusetts Teacher Association, 2011, p. 11).

The State of Massachusetts has put practices in place for districts to report to the Department of Education regarding individual educator evaluation data, including performance ratings on each standard and overall. This information will not be made public on an individual level although aggregate data that do not individually identify educators may be made public (MDESE, 2012).

The changes found in the 2011 Massachusetts state code were in part as a result of the recommendations from the Massachusetts Task Force on the Evaluation of Teachers and Administrators (MDESE, 2011). The task force consisted of over 40 educators including a variety of perspective such as teachers, building administrators, district administrators, university personnel, and state-level educators. The task force cited the following conclusions in their recommendations to their State Board of Education:

- Rarely includes student outcomes as a factor in evaluation
- Often fails to differentiate meaningfully between levels of educator effectiveness
- Fails to identify variation in effectiveness within schools and districts
- Rarely singles out excellence among educators
- Does not address issues of capacity, or “do-ability”
- Fails to calibrate ratings, allowing inconsistent practices across the state
- Fails to ensure educator input or continuous improvement

- Is often under-resourced or not taken seriously (MDESE, 2011, p. 5).

The task force further concluded that past poor evaluation practices have resulted in missed opportunities for improvements in leading, teaching, and learning. The task force also proposed the four rating categories that are now used by the State of Massachusetts.

The State of Massachusetts determined that all evaluations will be based on a 5-step cycle. This cycle is illustrated in Table 7.

Table 7

*Massachusetts Teacher Evaluation Cycle*

Step	Description
Self-Assessment	Propose goals for Educator Plan alone and in teams. Reflect and assess professional practice, analyze learning, growth, and achievement of their students.
Analysis, Goal Setting, and Plan Development	Educators and evaluators meet to review self-assessments. Jointly analyze students' learning and develop goals and plan that cover practice and student learning.
Implementation of the Plan	Educators implement action steps in plan and engage in professional development and support. Educator and evaluator collect evidence to inform progress.
Formative Assessment/Evaluation	Educator and evaluator review progress towards goals. Evaluator issues formative performance ratings.
Summative Evaluation	Evaluator assesses educator against standards of student learning and professional practice goals. Evaluator determines overall summative rating using 4-point rating scale. (and student learning rating on 3-point scale when applicable)

*Note.* Adapted from Massachusetts Department of Elementary and Secondary Education (2011). *Overview of the new Massachusetts educator evaluation framework*. Retrieved from <http://www.doe.mass.edu/eval/101511Overview.pdf>

The *State of the States* report was critical of the adopted regulations of Massachusetts. The report stated that the student learning measure “leaves too many details and too much discretion to individual evaluators to choose student achievement measures and make decisions about the adequacy of growth attained by individual teachers” (NCTQ, 2011, p. 10). The report cites that Massachusetts removed language that required measures to be a “significant” factor in educator evaluations. The report also cited that the large advisory committee of 45 people prevented consensus on reform and the committee was “seeming to back-pedal from more ambitious and rigorous expectations” (NCTQ, 2011, p. 34). However, as a point of comparison, student learning as a component of teacher evaluation in Illinois will not even go into effect for teachers as part of the evaluation process until 2015-16. Therefore, the criticisms about Massachusetts’ student learning component should have little impact on this study in relation to the perceived impact of the expanded teacher evaluation ratings.

### **Teacher Evaluation Reform in Florida**

In 2010, Florida became one of 12 winners of federal *Race to the Top* grant program funding. In Florida’s first year of funding from RTT the state revised teacher and principal evaluations, began the transition to Common Core State Standards, and developed eight different implementation committees. Also, in March of 2011, the Student Success Act was passed by the Florida Legislature which supported many of the goals from the RTT application. Florida touted *Race to the Top* as “an opportunity to broaden and accelerate [their] reforms to boost teacher effectiveness and the achievement of nearly 2.7 million students” (Florida Department of Education, 2013. p. 1).

The Florida Department of Education stated that through *Race to the Top* their vision was of student-centered school environments with teachers engaged in peer collaboration, using data to improve teaching, with the support of instructional leaders who are effective, which will result in increased student achievement. To that end they stated that their *Race to the Top* theory of reform was that “highly-effective teachers and leaders are key factors in improved student achievement” (*Florida’s Race to the Top*, 2013. p. 1).

Florida first had the four-tier model for teacher evaluation appear in their state code in 2011. The ratings they assign to their teachers are “highly effective,” “effective,” “needs improvement” (or, for instructional personnel in the first three years of employment who need improvement, developing), and “unsatisfactory” (Florida Legislative Statutes, 2011). Prior to this Florida had no teacher evaluation rating distinctions in their state code (Florida Legislative Statutes, 2010). The 2011-12 school year was also when each school district in Florida was expected to use student-learning growth in their evaluation formulas (NCTQ, 2011, p. 46).

As of 2012 the Marzano based Florida model teacher evaluation instrument was used in 44% of Florida school districts. The Danielson model was used in 20% of Florida school districts. The remaining districts used models that incorporated elements of one or both of the two models (Florida Department of Education, 2012, p. 2).

Florida bases their evaluation measures on student growth and also the four domains of classroom strategies and behaviors, planning and preparation, reflections on teaching, and collegiality and professionalism. Also, unique to Florida is that “parents must have an opportunity for input on teacher performance ratings” (NCTQ, 2011, p. 46).

Florida eliminated their tenure policies and now bases their contracts on performance in the classroom. The state has realigned their contract practices (see Table 8) so that now “To be awarded or renew an annual contract, a teacher must not have received any of the following evaluation ratings: two consecutive annual performance evaluation ratings of unsatisfactory within a three-year period; or three consecutive annual performance evaluation ratings of needs improvement or a combination of needs improvement and unsatisfactory” (NCTQ, 2011, p. 25).

Table 8

*Florida Poor Teacher Evaluation Consequences*

Policy for Assisting Teachers Who Receive Poor Evaluations	Teachers are Eligible for Dismissal Based on Poor Evaluations	Policy for Dismissing Ineffective Teachers
When a teacher receives an unsatisfactory evaluation the evaluator must make recommendations as to specific areas of unsatisfactory performance and provide assistance in helping to correct deficiencies within a prescribed period of time.	Yes.	Ensures that teacher ineffectiveness is grounds for dismissal. All new teachers are placed on annual contracts and the state requires that such contracts are not renewed if a teacher’s performance is unsatisfactory. An annual contract may not be awarded if the teacher has received two consecutive annual performance evaluation ratings of unsatisfactory, or two annual performance ratings of unsatisfactory within a three-year period, or three consecutive annual performance evaluation ratings of needs improvement or a combination of needs improvement and unsatisfactory.

*Note.* Adapted from National Council on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Retrieved from the National Council on Teacher Quality website: [www.nctq.org/p/publications/docs/nctq\\_stateOfTheStates.pdf](http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf)

Florida is continuing to develop and fine tune their new teacher evaluation practices and procedures. This includes a shift to requiring that teacher compensation

include teacher performance beginning in 2014. In this new model, “A teacher determined to be highly effective will receive a salary increase that must be greater than the highest annual salary adjustment available to that individual through any other salary schedule adopted by the school district” and an “effective” teacher will see a salary increase between 50 and 75 percent of the increase given to the highly effective teacher” (NCTQ, 2011, p. 26). Although Illinois has not proposed similar shifts in tying performance ratings to compensation the overall concept of expanding the teacher evaluation ratings is similar and can serve as a point of comparison in order to inform Illinois principals regarding the perceived impact of the expanded performance ratings. Table 9 illustrates the comparison of the three states on key teacher evaluation categories.

Table 9

*Comparison of Illinois, Massachusetts, and Florida Teacher Evaluation Systems*

Commonalities Between All Three States	Characteristics Unique to Illinois	Characteristics Unique to Massachusetts	Characteristics Unique to Florida
Teacher evaluation rating categories are comprised of four ratings.	Implemented expanded rating system in 2012-13.	Implemented expanded rating system in 2011-12.	Implemented expanded rating system in 2011-12.
Default evaluation models have connections to Charlotte Danielson’s Framework for Teaching.	Expanded from 3 to 4 rating categories.	No previous state guidance on evaluation rating number or categories.	No teacher evaluation rating distinctions in their previous state code.
Received funding from Race to the Top (FL, MA 2010, IL 2011).	Rating categories: -Excellent -Proficient -Needs Improvement -Unsatisfactory	Rating categories: -Exemplary -Proficient -Needs Improvement -Unsatisfactory	Rating categories: -Highly Effective -Effective -Needs Improvement -Unsatisfactory
Evaluation systems incorporate student growth components in their implementation.	Default state system based on Charlotte Danielson rubric.	Default state system based on Charlotte Danielson and Kim Marshall rubrics.	Default state system based on Charlotte Danielson and Robert Marzano rubrics.

### Summary

In conclusion, the goal of this literature review was to present current research to further inform and provide a context for the perceived impact of teacher performance ratings on the teacher evaluation process. This was first done by exploring a brief historical overview of teacher evaluation, analyzing teacher evaluation through the lenses of teacher recognition, teacher effectiveness, teacher growth, teacher remediation, and teacher dismissal. An analysis of the relationship between principal and teacher in the evaluation process, the courageous and strategic conversations that should take place in relation to teacher evaluation, and the perceived impact of the teacher evaluator was studied along with the perceived impact of evaluation on school culture and climate and how the evaluator can promote social justice through effective evaluation. Current trends and impacts on teacher evaluation were explored with specific emphasis on *Race to the Top* and Charlotte Danielson's *Framework for Effective Teaching*. Finally, teacher evaluation reform in Illinois, Massachusetts, and Florida was reviewed to complete the literature review.

## CHAPTER III

### METHODOLOGY

#### **Introduction**

The research methodology for the study was a cross sectional survey with quantitative and qualitative data. It was cross sectional because it was a snapshot in time of what people think or believe (Merriam, 2009). This resulted in a case study bounded by the states of Massachusetts and Florida. The study surveyed all K-12 public school principals in these two states. The goal of the study was to compare and contrast the perceptions and experiences of the principals in Massachusetts and Florida in order to inform Illinois principals about teacher rating usage and effectiveness on a number of levels between the models in Massachusetts and Florida and the new Illinois evaluation model.

The survey itself was built around the conceptual framework of Charlotte Danielson's four domains for effective teaching. The goal of all questions included in the survey was to support the overarching research questions which were to explore and measure the perceptions of principals concerning teacher performance ratings on teacher growth and effectiveness. Stated more specifically the survey intended to explore and measure the following from the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

The vision of implementation for this study was to provide an electronic opportunity for principals to participate. The survey was generated using Survey Monkey® due to the security and confidentiality it provided. Email addresses for K-12 principals in Massachusetts and Florida were obtained through information collected on

the states' respective department of education websites. The survey (see Appendix B) was sent out via email to each K-12 public school principal in the states of Massachusetts and Florida and was accompanied by a consent letter (see Appendix A) making them aware of the study and the request to participate. This survey was sent out two more times via email over the course of a three week period with a follow-up request for participation (see Appendices C and D).

The survey consisted of Likert scale and open-ended questions intended to collect information on the principals' perceptions of the impact of expanded teacher evaluation performance ratings on teacher recognition, effectiveness, growth, remediation, and dismissal. Demographic information was also collected so as to further analyze the data upon survey completion.

### **Research Design**

The research methodology of a cross sectional survey with quantitative and qualitative data was chosen for this study because the researcher intended to collect a significant quantity of data from principals in the states of Massachusetts and Florida, based on their present perceptions, in order to immediately inform Illinois principals as they follow a similar path to the principals in these two states. Timeliness of data collection was critical given that the goal was to inform Illinois based on perceptions of Massachusetts and Florida. A longitudinal study would not be as advantageous as a cross-sectional study that collects data at only one point in time instead of collecting over time. The goal of this design was to explore potential causal relationships between an expanded four-tier rating system and its perceived impact on teacher recognition, effectiveness, growth, remediation, and dismissal.

A qualitative approach was chosen as it would be difficult to manipulate variables in any way to study this topic with an experimental quantitative design. Instead, the study aimed to take advantage of variables that were already manipulated naturally which made the study qualitative as “the research seeks to establish the meaning of a phenomenon from the views of participants. This means identifying a culture-sharing group and studying how it develops shared patterns of behavior over time” (Creswell, 2009, p. 16). In contrast, a quantitative experimental design would measure variables both before and after the experimental treatment. A mixed methods design would also differ from this research design as it would follow the survey and data analysis with a second phase that would follow an interview approach to dig deeper into the results of the survey with detailed reviews from the participants which this study did not do.

This study would be considered a “natural experiment” as it described “a naturally-occurring contrast between a treatment and a comparison condition” (Shadish et al., 2002, p. 17). There could be no manipulation of variables in this research and instead the study analyzed the manipulation in variables that naturally occurred as the legislation in the states of Massachusetts and Florida mandated that principals change the rating system by which they evaluate teachers. The comparison condition in this study was the teacher evaluation rating system in the states of Massachusetts and Florida prior to 2011. The treatment condition was the four-tier teacher evaluation system that followed in 2011-12 for these two states. These conditions mirrored those that occurred in Illinois one year later and therefore the perceptions of principals in the states of Massachusetts and Florida served to inform the principals in Illinois who are following in these footsteps.

The form of data collection was an online survey questionnaire through a web-based administration. The choice of a survey for the research design was because the survey “provides a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population” (Creswell, 2009, p. 145). The survey was the ideal method for this research as it was easy to give in the online format, it provided for a quick response rate, and it also allowed for a large sample size to collect data from in comparison to personal interviews. This research method also allowed for responses to be anonymous, economical for the researcher, and convenient for both the researcher and participant.

Marshall and Rossman (2011) captured the rationale for making this design an online survey by stating “one major advantage in using the Internet to gather data is that one’s sample can quite literally be global” (p. 181). This was a critical attribute to the decision to make this research method based on an online survey given the distant locations of the populations that were surveyed in Florida and Massachusetts in comparison to the researcher’s locale in the Midwest.

### **Participants**

The participants for this study were practicing K-12 public school principals in the states of Florida and Massachusetts that had active email addresses posted on their respective state department of education websites in March of 2013, Massachusetts (n=1,854) and Florida (n=4,533). The survey was designed so that beginning principals that had only operated under their respective state’s current teacher evaluation system could be separated from those principals that had operated under the current teacher evaluation system and the prior teacher evaluation system. Given that non-public school

principals were not subject to all of the statutes related to teacher evaluation they were excluded from the study's participant pool.

The participants were surveyed in a single stage as opposed to multiple stages as the researcher had access to contact information from all of the participants and could sample them all directly at one time. In essence, the research was a random sample as participants were contacted through a blanket survey and their data was collected based on their desire to participate as opposed to any other criteria that could have narrowed down the selection pool of eligible participants. By the same token the participants were not stratified out for selection to participate in any way beyond their identification as a K-12 public school principal before they were contacted. No further controls were implemented such as gender, age, ethnicity, etc. However, these demographic data points were used in the subsequent analysis.

Email addresses for K-12 principals in Massachusetts and Florida were obtained through information collected on the states' respective department of education websites. The survey was sent out via email to each K-12 public school principal in the states of Massachusetts and Florida. The survey was accompanied by a consent letter making them aware of the study and the request to participate. This survey was sent out two more times via email over the course of a three week period with a reminder follow-up request for participation. The number of participants that were surveyed was based on the number of email addresses obtained in the two states. All emails provided were sent a survey and request to participate in the study.

### **Procedures for Data Collection**

This survey was conducted online using email addresses obtained in both Massachusetts and Florida. The survey instrument was not modified from any other existing survey instrument of which the researcher was aware. The survey was sent out via email to each K-12 public school principal in the states of Massachusetts and Florida and was accompanied by a consent letter making them aware of the study and the request to participate. The consent letter (see Appendix A) appeared on the first page of the online survey. More specifically, the consent letter of the survey provided the following information to the participants:

- Background of researcher
- Purpose of the study
- Rationale for participant selection
- Research areas
- Conceptual framework
- Participant instructions and voluntary nature
- Confidentiality and anonymity
- Contact information

The next page following the consent letter contained an embedded consent form which participants must have agreed to the terms of before they were allowed to continue on with the survey if they so chose.

Following the consent page of the survey was the survey itself, entitled “Survey on the Impact of Teacher Evaluation Ratings” (see Appendix B). The closed response

portion of the survey consisted of multiple choice and Likert scale questions. The survey first consisted of the following areas of focus with no open-ended questions:

- Demographic
- Perception Overview
- Teacher Recognition
- Teacher Effectiveness
- Teacher Growth
- Teacher Remediation
- Teacher Dismissal

These questions were followed by an open-ended question which provided the opportunity for the participant to further explain any answer from the survey. This concluded the participant's involvement in the survey. The survey was piloted with 14 Loyola University School of Education administration and supervision doctoral students in the fall of 2012. This allowed for the instrument to be further refined as the respondents took the survey and then provided feedback based on their experiences. This feedback included focused information around the validity of the survey in addition to effective formatting.

The survey was sent out two more times via email over the course of a three week period with reminder follow-up requests for participation if the participant had not already done so (see Appendices C and D). The first of these follow-up emails was sent approximately one week following the initial request for participation. The second follow-up email was sent approximately one week after the first follow-up with a

notification that the survey would close within the next two days. Once that deadline was reached the survey link was closed to participants so that data sorting and analysis could begin.

Participants that had evaluated teachers for two years or less were not able to complete the survey in its entirety as they did not have the necessary experience in order to compare and contrast the two teacher evaluation models that existed in their respective states. Their survey followed a different path based on their years of experience. This procedure was done so that these new principals were not asked questions that did not apply to them and also so that the researcher was not collecting information that would not be used. Their data was analyzed within a subgroup to determine if their perceptions possessed any meaningful information that would help to answer the overarching research questions. However, principals in Massachusetts and Florida who had been in their positions for more than two years and had therefore transitioned from their former rating system to an expanded four-tier rating system were given the full battery of questions.

Survey Monkey® was used as the instrument for online survey administration and data collection. According to the website's privacy policy page (Survey Monkey, 2012) this was a secure site that provided features to ensure safety and anonymity while administering the surveys and collecting data. The researcher was able to create custom templates for the survey which could then be emailed to participants to complete. According to Survey Monkey's® privacy policy Survey Monkey® would treat the surveys as private data owned by the user. They do not sell any information collected unless the user makes them public or if the company is compelled to do so by law.

Survey Monkey® allowed the user to upload lists of email addresses and the company only served as a custodian of that data. They do not sell email addresses and only use them as directed by the user.

Online survey instruments such as Survey Monkey® cannot actually perform the analysis and instead Microsoft Excel was used by the researcher to provide an in-depth analysis of the data. Survey Monkey® generated results which were then reported back to the researcher as descriptive statistics and information in graph form. This information was then downloaded into Microsoft Excel for further analysis by demographics and applied to the conceptual framework. Survey Monkey® was primarily used to compile the data and make it user-friendly in order to efficiently apply it to these other tools after collection.

Prior to each follow-up notification to complete the survey, the data was pulled off of Survey Monkey® and was backed up in a secure server that was independent of Survey Monkey® to provide additional data security.

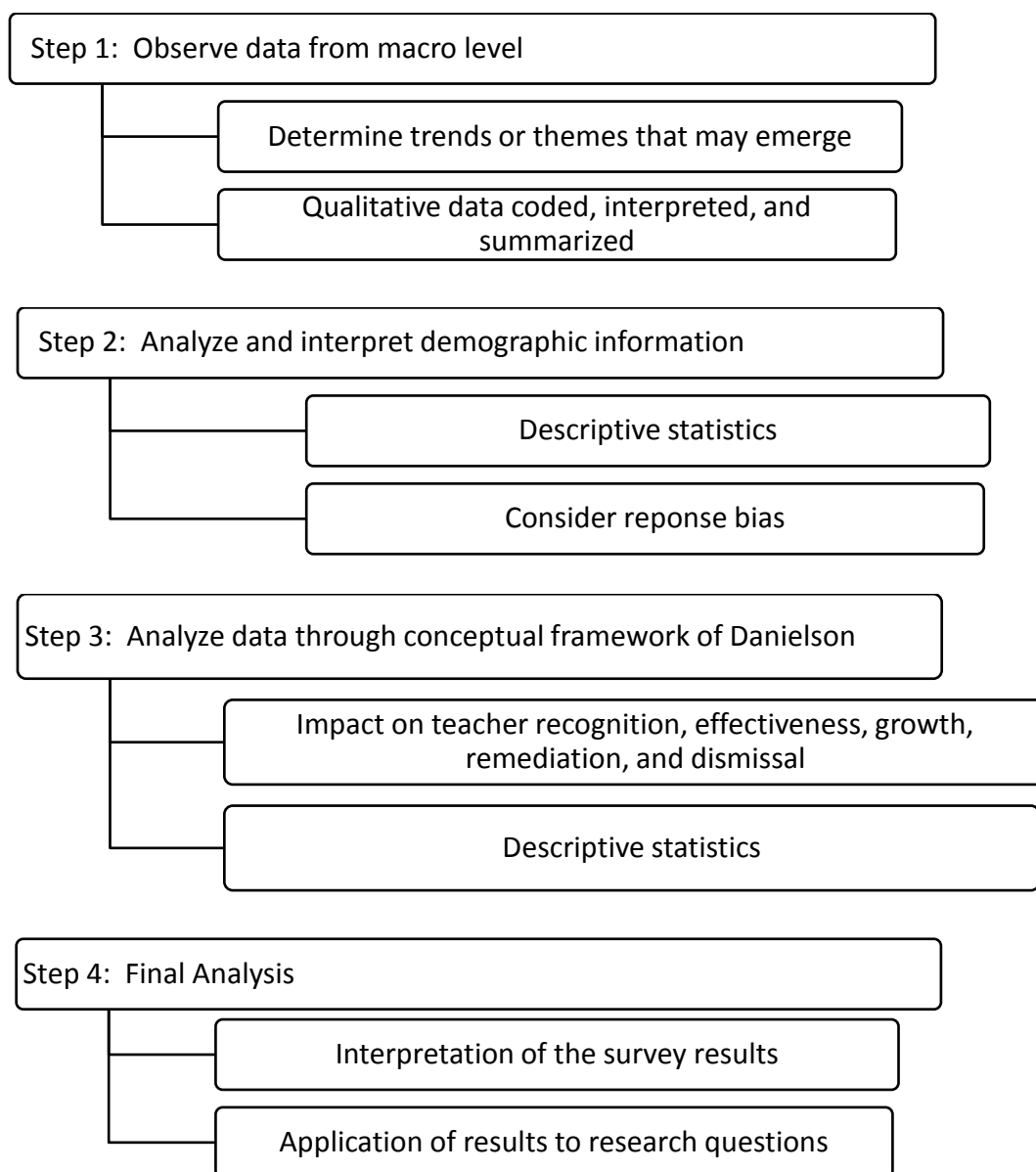
The goal of the three separate notifications of the online survey was to maximize the participant response rate. If an adequate number of responses were not obtained through the proposed survey administration detailed in this section then a time extension to the survey with additional notifications would have been considered.

Once the survey link window was closed then all of the quantitative and qualitative data was collected and compiled. The quantitative data was downloaded to Microsoft Excel for further analysis. The qualitative data was coded by the researcher so that it could be further analyzed.

### **Data Analysis**

The collection of data was an important step in informing the research but the analysis of these data was where the researcher found meaning to inform and answer the research questions. As stated by Merriam (2009), the qualitative design is emergent and “the process of data collection and analysis is recursive and dynamic” while the “analysis becomes more intensive as the study progresses and once all the data are in” (p. 169). Thomas A. Schwandt (2007) further defined the importance of data analysis by stating “what constitutes data depends upon one’s inquiry purposes and the questions one seeks to answer” and that conceptual schemes need to be generated to effectively analyze data and attempt to answer the research questions (p. 128).

The first step in the analysis of these data was to observe the data from a macro level to determine any overarching trends or themes in the responses that emerged. The second step in the data analysis was to focus on and interpret the demographic information collected. Third, the data was analyzed using the conceptual framework of Danielson to see how the principals perceived the four-tier evaluation system’s impact on teacher recognition, effectiveness, growth, remediation, and dismissal (see Figure 3).



*Figure 3. Data Analysis Procedures*

Data was organized so that it could be analyzed comparing the Danielson domains to the teacher performance rating perceived impacts on teacher recognition, effectiveness, growth, remediation, and dismissal as shown below:

Table 11

*Data Analysis Framework under Four-tier Teacher Evaluation System*

	Planning and Preparation	Classroom Environment	Instruction	Professional Responsibilities
Teacher Recognition				
Teacher Effectiveness				
Teacher Growth				
Teacher Remediation				
Teacher Dismissal				

Data was also collected to determine members of the participant pool that did not participate in the study and this data was reported out. Response bias was also considered to see if the responses of non-respondents could have substantially impacted the overall results. As Creswell (2009) defined it, “response bias is the effect of nonresponses on survey estimates” (p. 151).

The qualitative data from the open-ended questions were analyzed by the researcher to determine if any themes emerged from these data that could further inform the overarching research questions. These data were coded, interpreted, and summarized. As Marshall and Rossman (2011) suggested “using both the readings of the data, and the conceptual framework for indications, the researcher sees how the data function or nest in their context and what varieties appear and how frequently the different varieties appear” (p. 213). By creating categories and themes the open-ended data were coded which then

led to clusters emerging which were further analyzed. In doing so, this qualitative data provided further information to supplement the quantitative data collected in the survey and further informed the study.

In the end the data analysis provided an interpretation of the survey results. Through the use of tables, figures, and rich descriptions the data were thoroughly presented and analyzed. The next step was to apply the results back to the research questions and determine to what extent the research answered these questions. As Creswell (2009) suggested the interpretation should then lead to the researcher indicating what might explain why these results occurred and discussing the implications of these results for practice or for any future research on this topic.

### **Ethical Considerations**

Ethical considerations were taken into account when developing this study. The Loyola University Institutional Review Board (IRB) Online Survey Research site was consulted during the design of the study to ensure participation was voluntary and with informed consent, information was confidential, and possible risks were addressed (Loyola University Institutional Review Board, 2013).

The purpose of the survey and methods used were clearly delineated in the consent letter of the survey. The researcher ensured that participation in this study was voluntary and contained an informed consent component in the survey as well. The design for collection and display of data was designed to ensure that confidentiality of participants was maintained and information requested that would identify a specific individual was eliminated. There were no known possible risks to participants as a result of their participation in this study.

### **Minimization of Bias**

The researcher acknowledges that personal bias may exist as he has personally evaluated teachers in three different teacher evaluation rating systems. The first system was a binary system where teachers were either given the rating of “Meets Expectations” or “Does Not Meet Expectations.” The second system was a three-tiered system which consisted of “Excellent,” “Satisfactory,” and “Unsatisfactory.” The third system is the newly adopted Illinois four-tiered teacher evaluation rating system which identified teachers as either “Excellent,” “Proficient,” “Needs Improvement,” or “Unsatisfactory.”

The intent of this design was intentional in trying to minimize personal bias and other biases by being largely quantitative in nature which would diminish the opportunities for bias to impact the results of the research. This does not eliminate bias that could impact everything from the design of the questions, the selection of participants, and the interpretation of results but it should decrease its likelihood of doing so.

To further account for and attempt to minimize bias the researcher kept a researcher’s journal. This journal was kept electronically and contained regular entries during the data collection and analysis process. The entries chronicled the impressions, feelings, and ongoing interpretations during the research. By revisiting the journal following the data analysis the researcher did not discover any biases that may have existed during data analysis that could have a bearing on how the results were interpreted and reported.

The opportunity to receive and implement feedback occurred on multiple levels. The pilot survey group [see Validity and Reliability], the dissertation director, and the dissertation readers served to minimize bias as the research was conducted.

### **Validity and Reliability**

The importance of validity and reliability in the study were critical to the usefulness of the study and “can be approached through careful attention to a study’s conceptualization and the way in which the data are collected, analyzed, and interpreted, and the way in which the findings are presented” (Shadish et al., 2002, p. 210). Through the research design that was detailed throughout this chapter and the presentation of data, analysis, and discussion that followed in the forthcoming chapters, the intent was to provide evidence that this study does indeed succeed in being reliable and valid.

To further promote validity the survey was piloted with 14 Loyola University School of Education administration and supervision doctoral students in the fall of 2012. The respondents took the survey and then provided feedback based on their experiences. This feedback included focused information around the validity of the survey in measuring the intended outcomes indicated in the proposed research questions. The feedback also included advice on effective formatting to make the survey more efficient to take and easier to read.

Although there was no way to prove perfect reliability this study was designed to maximize reliability. Given the fixed set of survey questions that all participants received there was little opportunity to obtain different information with repeated attempts at the same study. Given Schwandt’s (2007) definition of reliability which stated that “an account is judged to be reliable if it is capable of being replicated by another inquirer” (p.

262) this study should prove to have a high level of reliability. Therefore, Illinois principals could use this study to learn important information about their new evaluation model by understanding principal perceptions from Massachusetts and Florida in regards to measuring teacher performance and the resulting perceived impact on teacher recognition, effectiveness, growth, remediation, and dismissal.

### **Limitations**

There are several identified limitations to this study:

- 1) This study only focused on two states. The researcher uncovered other states that fit the same profile of Massachusetts and Florida but these two states were chosen so as to place parameters around the study that were more reasonable for the researcher to accomplish. Other states may have generated data that was similar or different based upon their own history in reaching this point in their teacher evaluation system in addition to other potential variables.
- 2) The researcher used self-selection in the form of unique sampling. As described by Merriam (2009) a unique sample “is based on unique, atypical, perhaps rare attributes or occurrences of the phenomenon of interest” (p. 78). This uniqueness made the states of Massachusetts and Florida good candidates for further research as they fit the necessary profile of being one of the first to expand their teacher rating systems. Their uniqueness may also have led to limitations as their experiences may not reflect those of other states that adopt similar legislation after them but may have other variables, such as learning from the experiences of early adopters, impact how they perceive and utilize the expanded teacher evaluation ratings.

- 3) This study relies on the recollection of participants. Memory is prone to error and therefore the data collected was also prone to the same inaccuracies. The design of the study was to collect principal perceptions and these perceptions have the potential to differ from reality. A large sample size may have had the ability to minimize this limitation.
- 4) The researcher had no way to verify the information reported was accurate as it was collected anonymously.

Despite these limitations this study was important because it was still able to provide insight to inform Illinois principals' practice in the area of teacher evaluation. The data collected was able to provide information that identified distinct trends and themes from a significant population of principals in Massachusetts and Florida. The limitations should be considered but should not preclude the data gathered from being deemed significant in advancing the study of the impact of expanded teacher performance ratings.

### **Summary**

The research methodology for the study was a cross sectional survey with quantitative and qualitative data bounded by the states of Massachusetts and Florida. The study surveyed all K-12 public school principals in these two states. The survey itself was built around the conceptual framework of Charlotte Danielson's four domains for effective teaching. The goal of all questions included in the survey was to support the overarching research questions which were to explore and measure the perception of principals concerning expanded teacher performance ratings on teacher growth and effectiveness. The vision of implementation for this study was to provide an electronic

opportunity for principals to participate. The survey consisted of Likert scale and open-ended questions intended to collect data to be analyzed and interpreted in relation to the research questions.

## CHAPTER IV

### PRESENTATION OF DATA

#### **Introduction**

The proposed research questions attempt to focus on the various perceived impacts, intended and unintended, created by the state mandated expansion to a four-tier teacher performance evaluation rating model. In order to capture these perceived impacts the following research questions have been researched:

From the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?

- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

### **Review of the Survey Administration**

The researcher first attempted to compile the emails of K-12 public school principals in Florida and Massachusetts through the use of Freedom of Information Act (FOIA) requests. One state referred the researcher to their state department of education website. The other state never responded, however, the researcher was able to find a database of emails on their department of education website as well. There were 6,387 K-12 public school principal positions listed on the website. All K-12 public school principal email addresses were pulled from Florida and Massachusetts state websites as they were posted in June of 2013. When preparing the email addresses for distribution it was discovered that the actual number of email addresses was significantly lower as many of the listings were either website addresses instead of email addresses, duplicate addresses, were not listed as a valid email address structure, or had email addresses which were omitted entirely. Therefore the number of email addresses totaled 4,459 principals in Florida and Massachusetts.

The researcher intended to use Survey Monkey® to distribute the surveys via email but their user policy prohibited the distribution of unsolicited emails due to anti-spam laws. The researcher contacted Survey Monkey® to inquire about an educational exemption to the company's policy, but none existed. The researcher next attempted to use Gmail to send out the survey. In researching Gmail it was understood that the email server would send email to 500 recipients per day per account which was higher than similar email service providers. The researcher opened up nine different accounts and uploaded 500 email addresses into each account so that all requests for surveys could be sent on the same day.

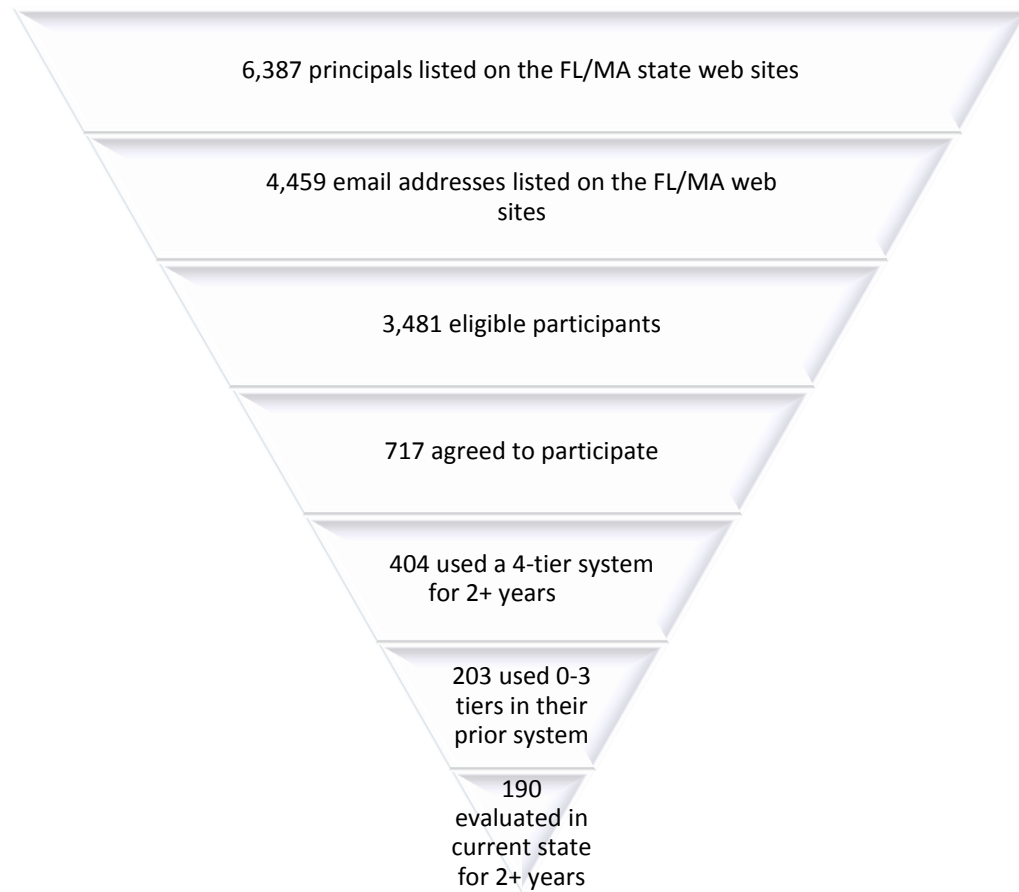
The first round of surveys was distributed through Gmail with a link to the study survey on Survey Monkey®. Over the next couple of days the response rate was low totaling less than 100 responses. The researcher hypothesized that emails were not all delivered by Gmail. The researcher revisited the Gmail accounts and discovered that the survey had been closed due to what Gmail cited as a "violation of their terms of use." The researcher attempted to determine what this violation was as the send limits of 500 emails per account were respected per the Gmail user policy. All attempts to call and email Gmail were not returned and instead inquiries were met with automated messages stating the accounts would not be reinstated.

After exploring several different options, the researcher concluded that the Loyola University Outlook account was a viable option to send out the follow-up survey reminder. The Loyola University account allowed an individual email to be sent to 500 recipients and had no limit to the number of emails that could be sent out in a day. The researcher sent out the follow-up request and eclipsed in one hour the number of

responses it took to collect in the first week. This led the researcher to conclude that the initial request did not reach all of its intended targets. The final follow-up email request was also sent through the Loyola University email account and was met with similar success in responses.

In total, 4,459 surveys were sent out via email to K-12 public school principals in Massachusetts and Florida (see Figure 4). Nine hundred and seventy-eight of these email invitations were removed from the sample because either the address was undeliverable, the participant responded that permission was needed from their district office to participate, the principal's out-of-office message indicated they were not replying to emails for the duration of the study, or the principal responded back that they were not an eligible candidate (i.e., new to position, did not adopt state model, non-traditional school, etc.). This reduced the sample size to 3,481 candidates.

When the survey link was closed there were 717 principals that agreed to participate in the study. This represented 20.6% of the sample size. The number of principals that evaluated teachers with a four-tier system for two or more years (Q8) was 404 principals which was 11.6% of the sample size. Also important was that the number of principals that had used a four-tier model for at least a year, and less than four tiers prior to their current model, was 399 principals, which was 11.5% of the total sample size.



*Figure 4.* Identification of the Survey Target Population

The target population for this study was principals that have evaluated teachers under a four-tier model for at least the past two years and have previously worked under a model with less than four tiers. This target population was selected because it was assumed that these principals would be able to inform the research from the perspective of having undergone an expanded rating system prior to Illinois which could therefore inform Illinois. In order to identify this target population the 717 participating principals were first sorted to include only those who had used the four tier model for two or more years which reduced the sample to 404 principals. Next, these data were sorted from this

sample of 404 principals to only include those that had used 0-3 performance tiers in their prior system (excluding those that previously used four or more tiers) which reduced the sample to 203 principals. Lastly, data were sorted to pull out any principals who had evaluated teachers for less than two years in their current state. This left a target population of 190 principals which represented 5.5% of the overall sample size (n=3,481) and 26.5% of responding principals (n=717). Another 201 principals previously used four tiers and continue to do so with their new respective state model which prevented another 28.0% of the responding principals (n=717) from meeting the target population criteria.

### **Data Presentation**

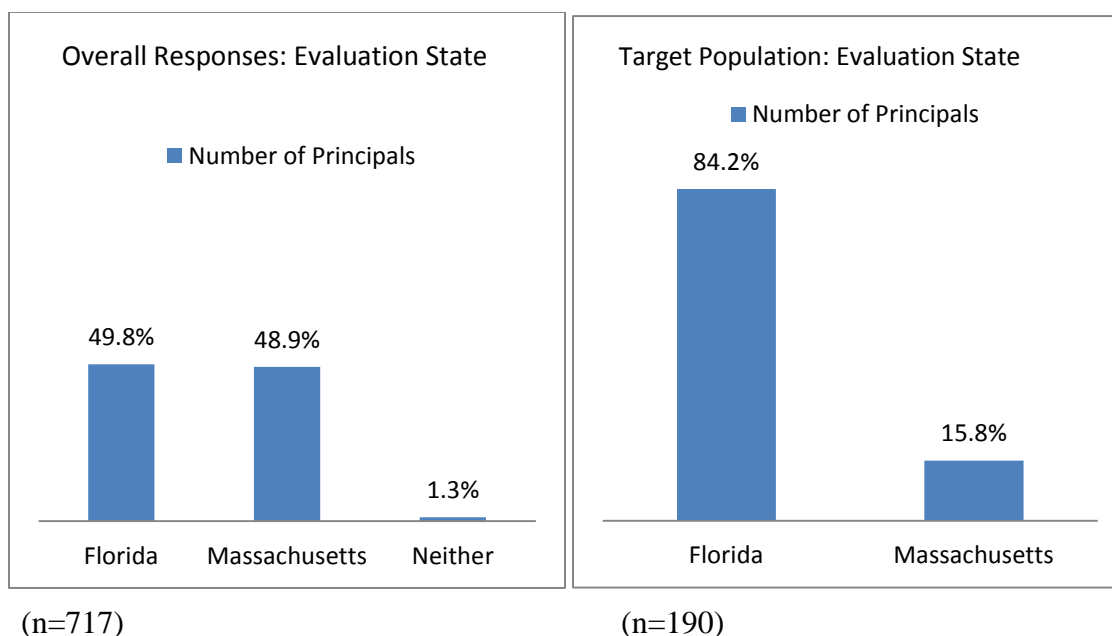
The data presentation is divided up into nine distinct areas. They are:

- 1) Demographic information;
- 2) Evaluation background, experiences, and perspective;
- 3) Teacher recognition;
- 4) Teacher effectiveness;
- 5) Teacher growth;
- 6) Teacher remediation;
- 7) Teacher dismissal;
- 8) Additional Comments;
- 9) Conceptual Framework.

### **Demographic Information**

As seen in Figure 5, there was a nearly even split in the number of principals that responded per state between Florida (n=355) and Massachusetts (n=348). Nine

principals reported that they did not evaluate teachers in either state in the past year. The number of principals that fit the profile of the target population for this study (see Figure 5) was not as even between Florida (n=160) and Massachusetts (n=30). This will be further discussed later in the chapter while analyzing Figure 11.



*Figure 5. State that Public School Principal Evaluated Teachers in the Past Year*

Overall, the principals surveyed had a varying background of educational experience (see Figure 6). The highest educational degree they obtained related to the field of education was significantly weighted towards a Master's degree (n=466). The other levels of education included: an advanced degree beyond a Master's (n=176), a Bachelor's degree (n=7), and those that reported "other" (n=56). In regards to the educational levels of the target population the trends were similar with over two times the number of MA/MS degree holders (n=128) in comparison to EdS/EdD/PhD degree holders (n=57).

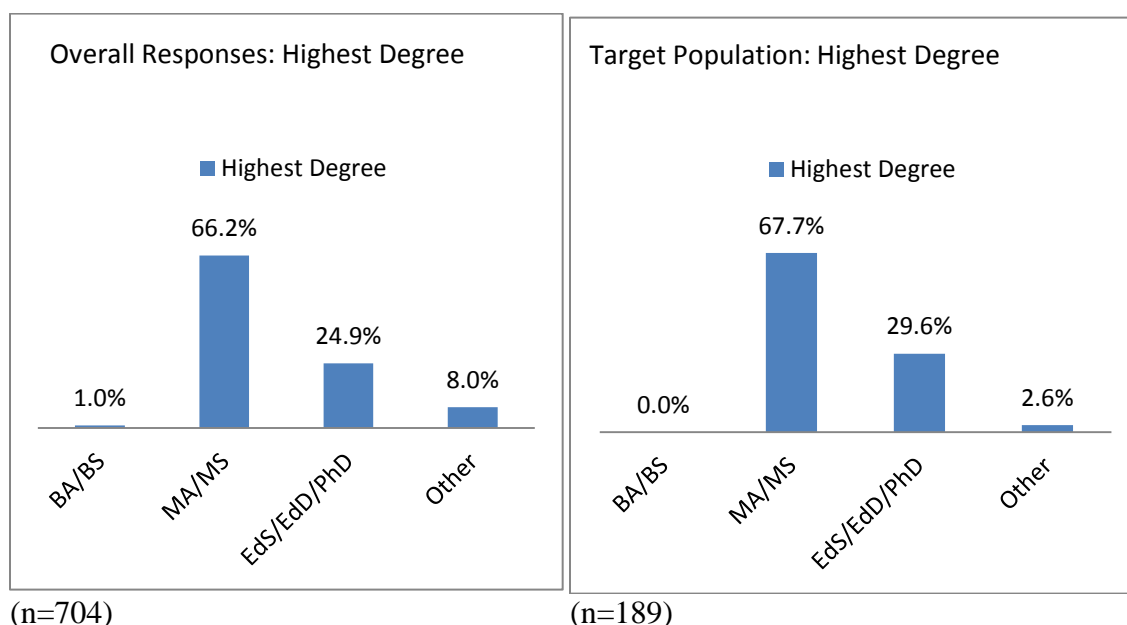


Figure 6. Educational Background

Figure 7 illustrates that the principals under 30 years of age were the smallest group ( $n=2$ ) followed by those between 30 and 39 years ( $n=80$ ). Principals between 40 and 49 years ( $n=241$ ) and between 50 and 59 years ( $n=274$ ) represented the two largest subgroups and then the number dropped back down for those that responded they were over 60 years ( $n=104$ ). The age range of the target population (see Figure 7) followed a similar pattern and never deviated by more than 2.4% for any subgroup in comparison to the overall responses.

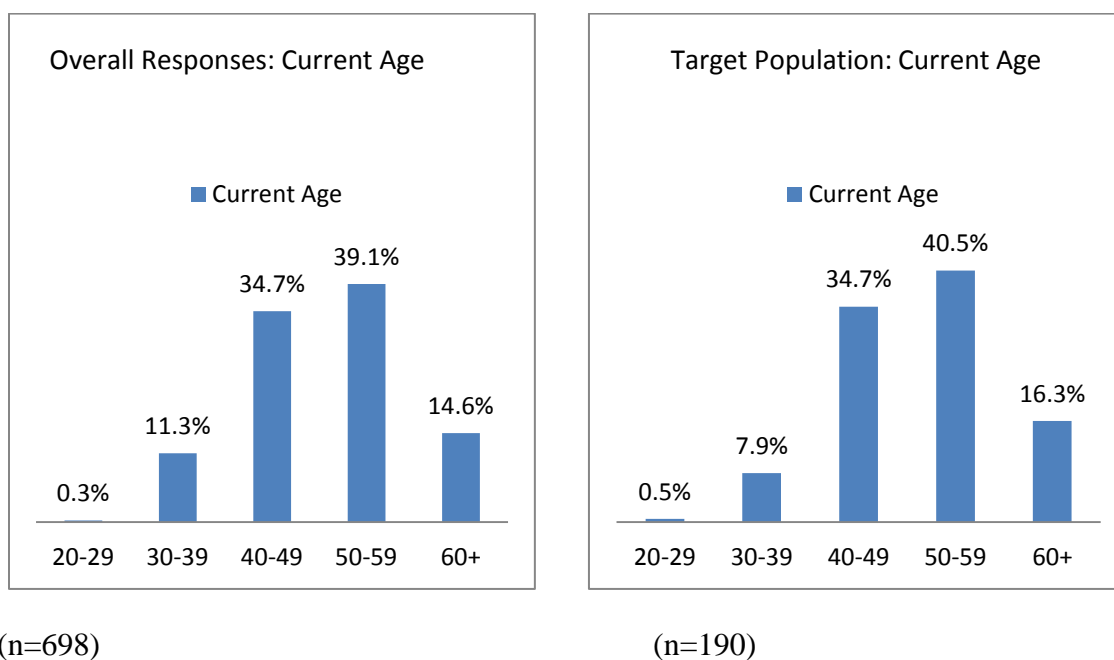
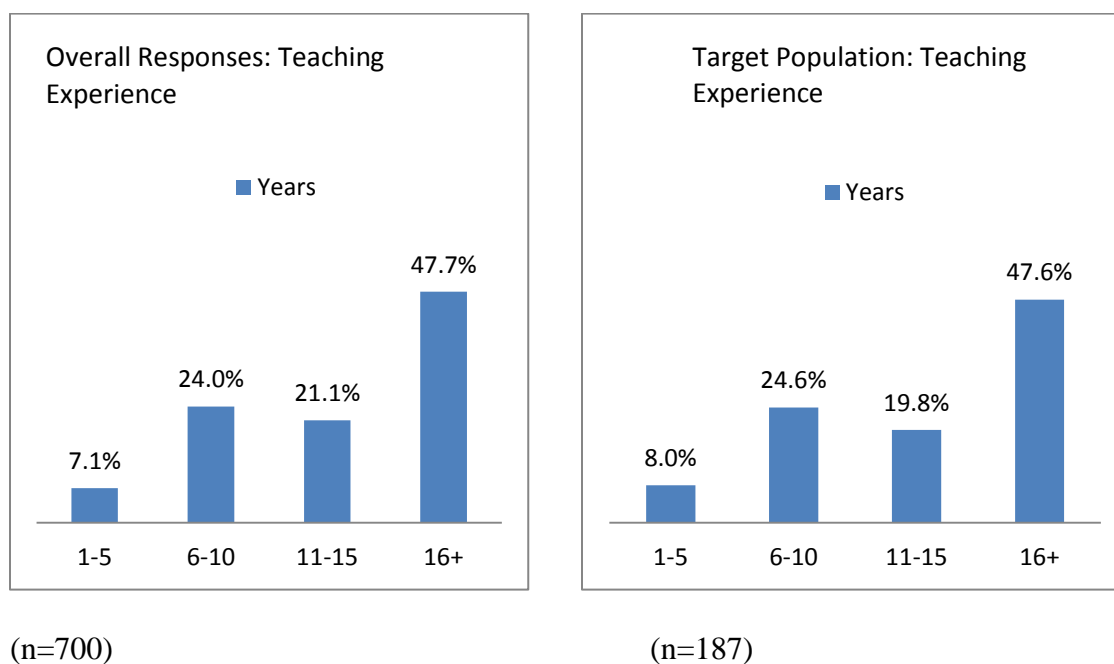


Figure 7. Current Age of Respondent

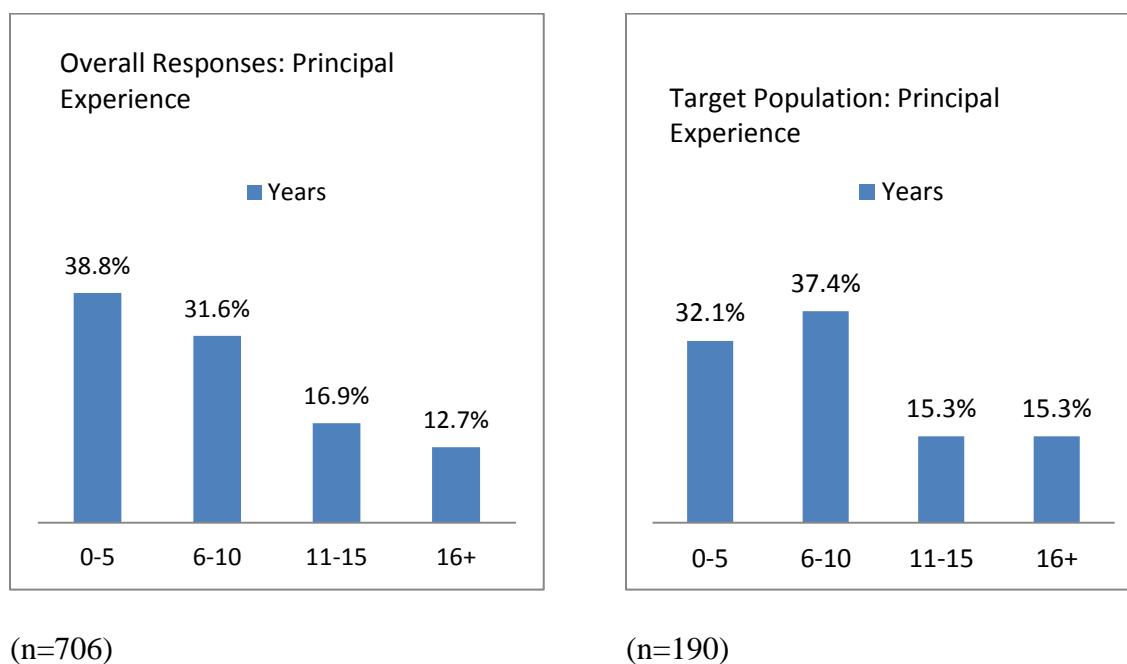
### Evaluation Background, Experience, and Perspective

Overall, the principals predominantly had 16 or more years of teaching experience (n=334) (see Figure 8). An almost equal number of principals fit into the other subgroups when combining principals reporting 1-5 years (n=50), principals reporting 6-10 years (n=169), and principals reporting 11-15 years (n=148). The years of teaching experience for the target population once again mirrored that of the overall sample size with no more than a 1.3% fluctuation for any subgroup (see Figure 8).



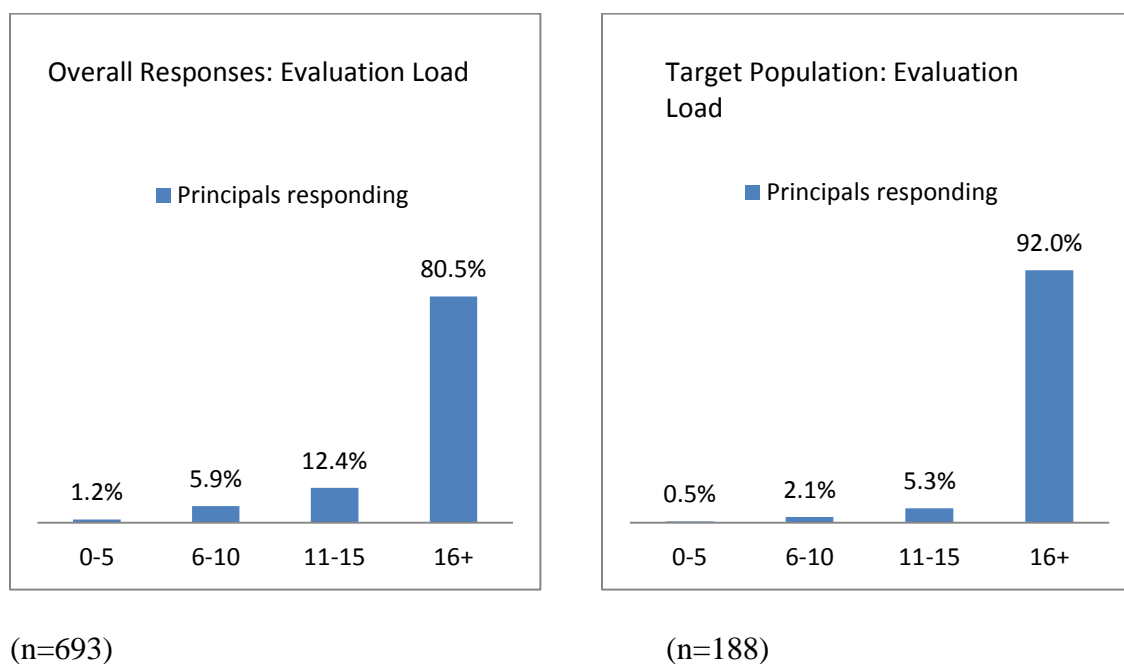
*Figure 8. Years of Teaching Experience*

As displayed in Figure 9, the number of years the respondents reported they have been a principal decreased by subgroup with each five year increment for the overall sample size. The largest group was principals who indicated they have served in this capacity for 5 years or less ( $n=274$ ), followed by those serving 6-10 years ( $n=223$ ), those serving 11-15 years ( $n=119$ ), and principals serving 16 years or more ( $n=90$ ). The target population reflected that the experience of principals fluctuated slightly with the largest subgroup appearing at 6-10 years of experience (see Figure 9).



*Figure 9. Years of Principal Experience*

The approximate number of teachers that the responding principals evaluated was overwhelming 16 or more per principal (n=558) as noted in Figure 10. Coming in at significantly lower totals were principals that indicated they evaluated 11-15 teachers (n=86), principals indicating that they evaluated 6-10 teachers (n=41), and principals indicating 0-5 teachers (n=8). The target population showed an even greater percentage of principals evaluating 16 or more teachers at a rate that was 11.5% higher than the overall response rate (see Figure 10).

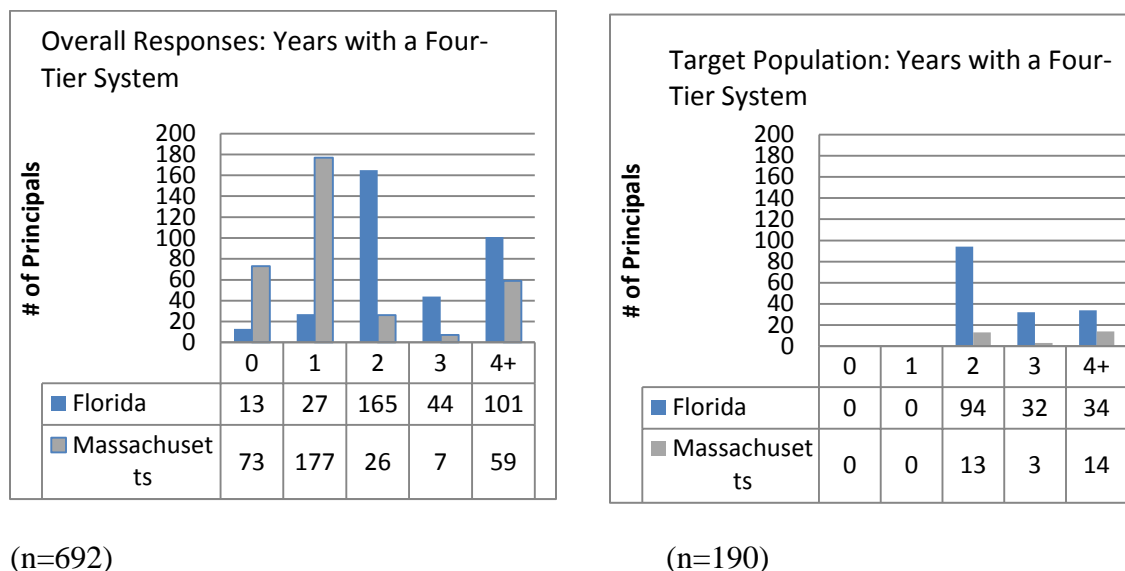


*Figure 10. Approximate Number of Teacher Evaluated Per Year by Principal*

The total number of years that the principals evaluated teachers using a performance rating system with four years explains why the survey target population is weighted towards Florida principals. Figure 11 represented that Florida had their largest surge of principals using a four-tier performance rating system two years ago (n=165) which indicated that this subgroup had used the system for two years. Massachusetts showed a similar surge one year later (n=177) which indicated this subgroup had used a four-tier system for one year. There was a population of principals in both Florida (n=13) and Massachusetts (n=73) that indicated they had yet to use a four-tier system. On the flip side there were principals in Florida (n=145) and Massachusetts (n=66) that indicated they had used a four-tier system for three or more years.

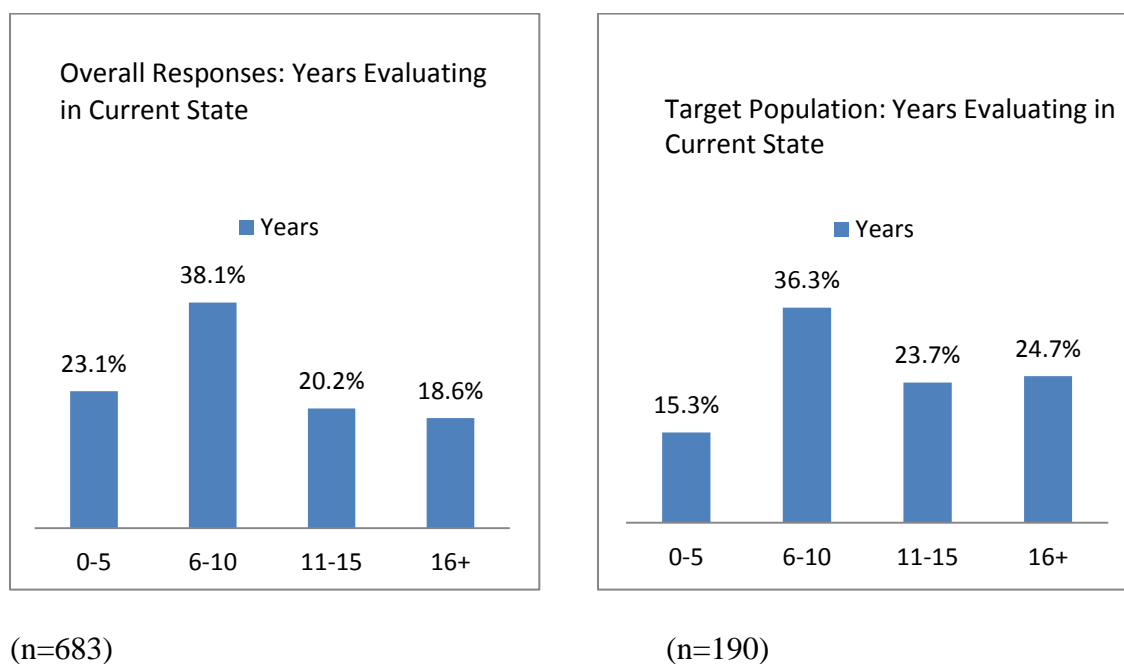
The target population saw far more Florida principals (n=160) in comparison to Massachusetts principals (n=30) that had utilized a four-tier evaluation system. This was

due in part to the fact that the target population removed any respondents that had not used a four-tier system for two years or more (see Figure 11).



*Figure 11. Total Years Principal has Evaluated Teachers Using a Four-Tier Performance Rating System*

The vast majority of respondents that evaluated teachers in their state were for well over five years, as seen in Figure 12. The largest subgroup was principals who had evaluated teachers in their respective state for 6-10 years (n=260), followed by 11-15 years (n=138), and 16 or more years (n=127). Those principals who evaluated teachers for 0-5 years (n=158) made up the second largest subgroup. The target population displayed similar trends although an even higher percentage of principals fit into the category of more than five years in the target population (74.7%) compared to the overall respondent group (76.9%), as seen in Figure 12.

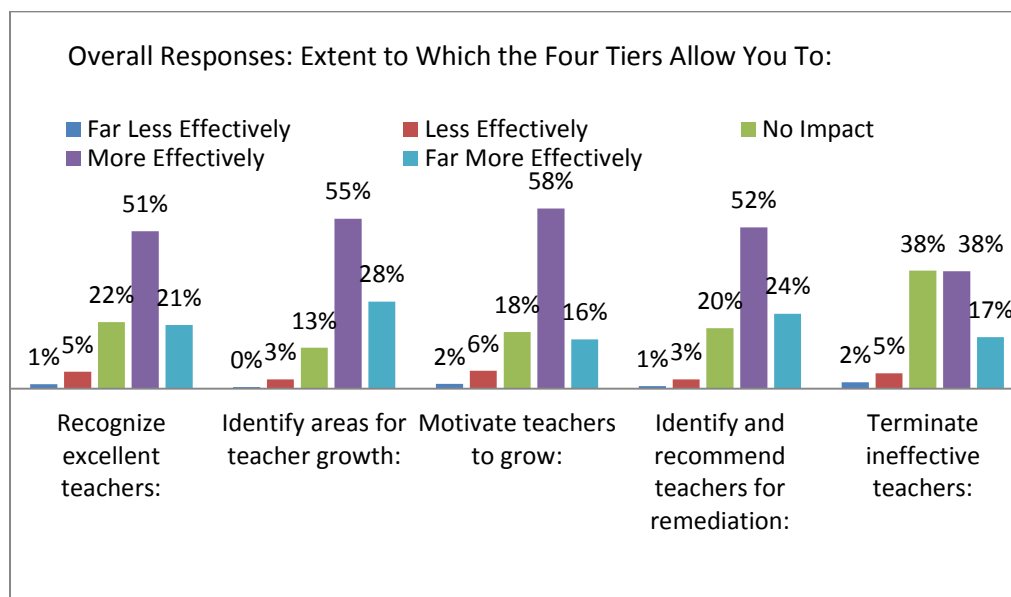


*Figure 12. Years Evaluating Teachers in Principals' Current State*

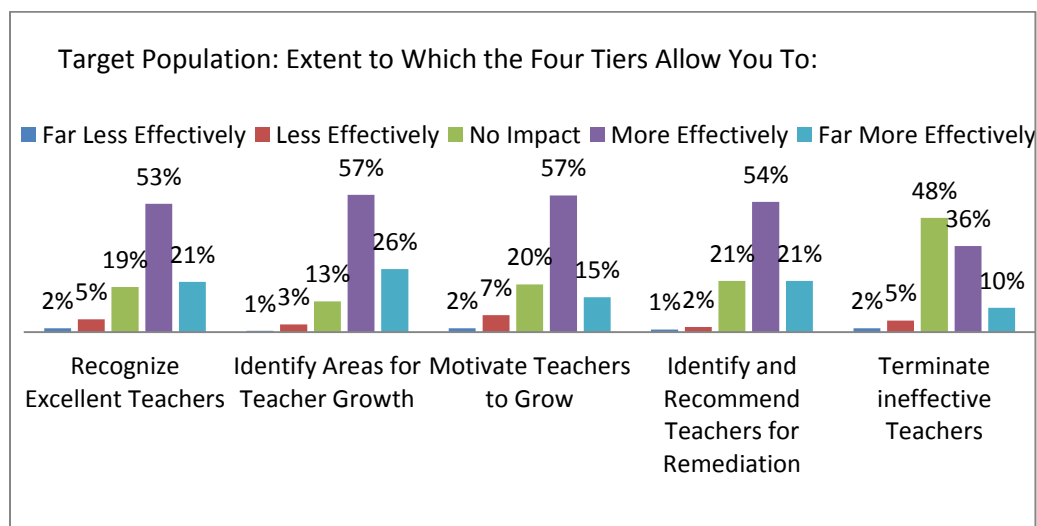
When principals were asked to compare their prior system to the four-tier rating system in several areas the majority of responding principals reported that the four-tier performance rating system was more effective or far more effective, as seen in Figure 13. This was true in respect to recognizing excellent teachers (n=459, 72%), identifying areas for teacher growth (n=535, 83%), motivating teachers to grow (n=475, 74%), and identifying and recommending teachers for remediation (n=482, 76%). In regards to terminating ineffective teachers the trend changed and there was an even split between principals that believed there was no impact (n=245) and principals who believed their current system was more effective (n=244). The principals were permitted to select multiple categories which were why the total responses exceed 100%.

In comparison, the target population that had lived the extended four-tier model for two or more years showed similar trends, as seen in Figure 13. The exception was the

ability to more effectively or far more effectively terminate ineffective teachers for the target population (46%). This target population rate was lower than the rate for the overall responses (55%).



(n=642)



(n=188)

*Figure 13.* Comparison of the Four-Tier Rating System to Principals' Prior Rating System

When principals were asked how many performance ratings they would prefer, Figure 14, the majority of responses reflected the four tier model (n=394). The next highest total was principals who preferred five performance rating tiers (n=175), followed by three tiers (n=55), two tiers (n=35), 6 or more tiers (n=5), and zero tiers (n=2). As seen in Figure 14, the target population once again showed similar trends although 7.8% fewer principals in the target population preferred a four-tier system to evaluate teachers.

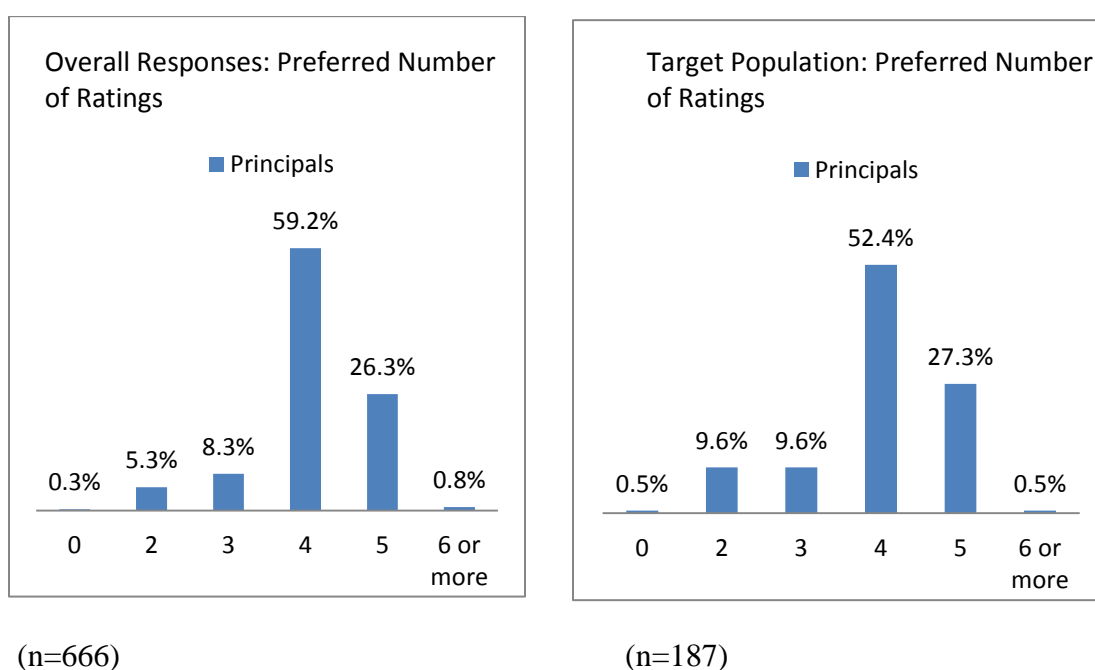


Figure 14. Number of Performance Ratings Preferred for Teacher Evaluation

### Teacher Recognition

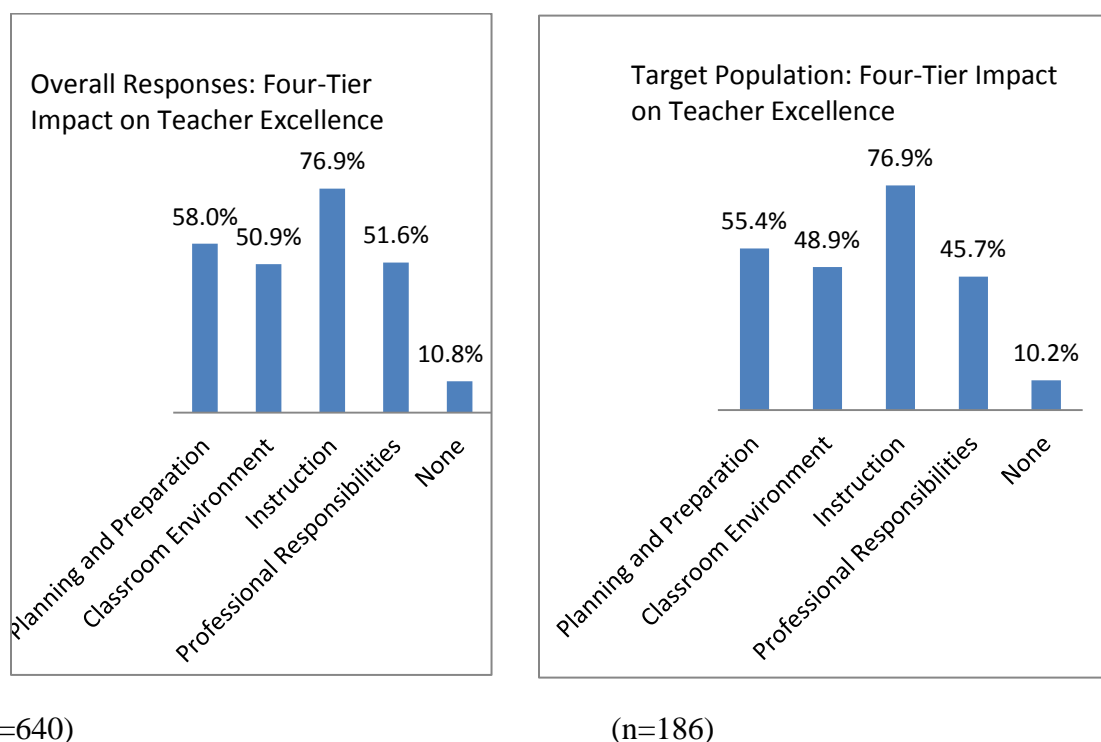
When principals were asked how the expanded four-tier teacher rating system could more effectively recognize teacher excellence compared to their prior teacher rating system in the domains of planning and preparation, classroom environment, instruction, and professional responsibilities the overall responses were over 50% for all

domains. The principals were permitted to select multiple categories which were why the total responses exceed 100%.

Figure 15 illustrates that the highest area of impact was teacher instruction (n=492, 76.9%) which was followed by planning and preparation (n=371, 58.0%), professional responsibilities (n=330, 51.6%), and classroom environment (n=326, 50.9%). Some principals felt that none of the domains effectively impacted the recognition of teacher excellence (n=69, 10.8%). The target population was slightly lower for every subgroup except planning and preparation which was equal. This target population still identified instruction within a four-tier system as the domain that most effectively recognizes teacher excellence in a four-tier system (n=143, 76.9%) (see Figure 15). Professional responsibilities showed the largest decrease between the overall responses and the target population with a 5.9% decrease (n=85, 45.7%).

The principals were asked to provide a written response regarding what they believed were the intended impacts of the four-tier rating system in relation to teacher recognition in their state (see Figure 16). After coding the responses from the target population the most frequent response related to the promotion of teacher growth at 32.5% as detailed by the following examples:

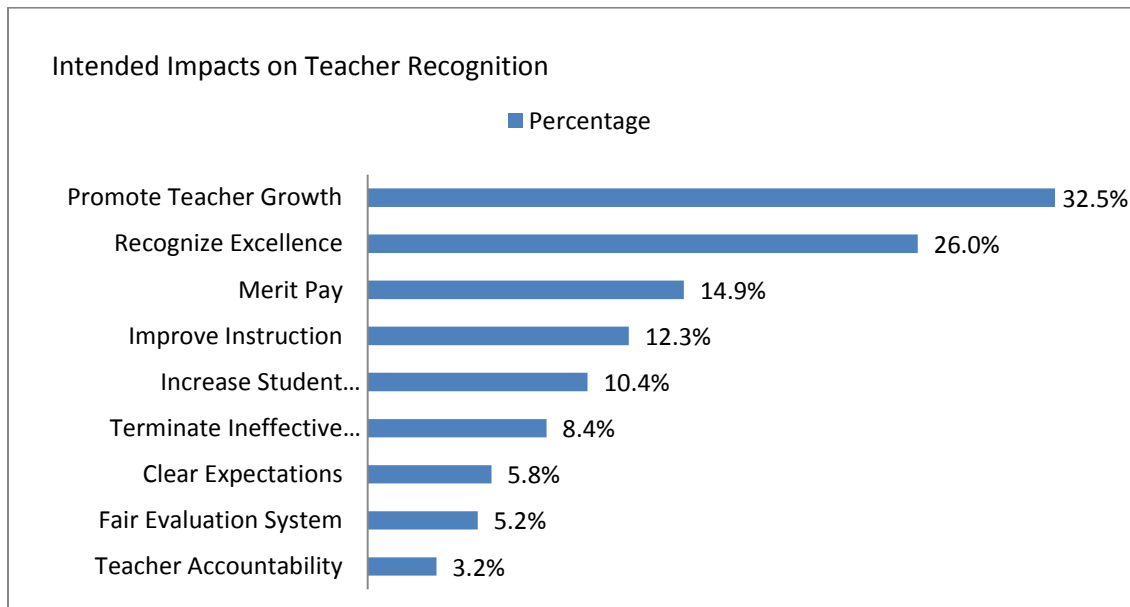
- “The intended impact is to more effectively identify exemplary teacher performance and to use teachers identified as exemplary as models for general staff development and strengthening of overall teacher performance within a building.”



*Figure 15.* Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Recognize Teacher Excellence

- “To truly differentiate teacher performance and impact in order to optimally communicate overall strengths as well as opportunities for improvement across various performance standards.”
- “Raise the bar of what makes an exceptional teacher and determine calibration for recognition.”

This theme was followed by recognizing excellence, establishing merit pay, improving instruction, increasing student achievement, terminating ineffective teachers, providing clear evaluation expectations, providing a fair teacher evaluation system, and teacher accountability. The principals were permitted to select multiple categories which was why the total responses exceed 100%.



(n=154)

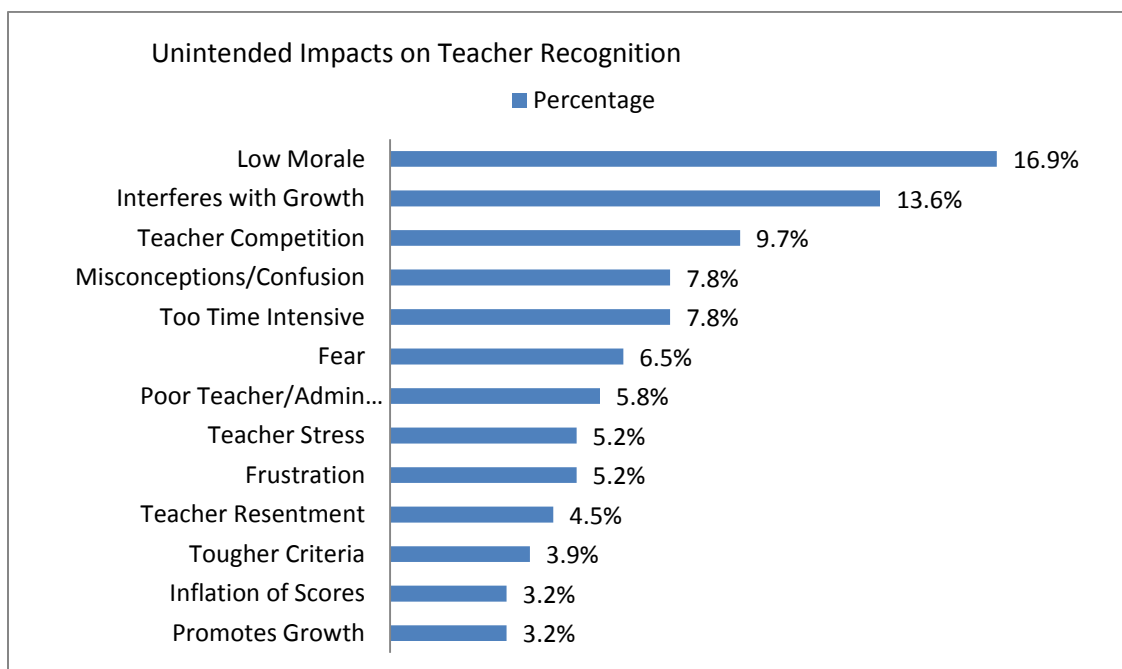
*Figure 16.* The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Recognition

The principals were next asked to provide extended responses regarding what they believed were the unintended impacts of the four-tier rating system in relation to teacher recognition in their state (see Figure 17). This response coding from the target population revealed the most frequent response was related to low teacher morale at 16.9% as detailed in the following examples:

- “Teacher morale has been very low with the new four tier rating system as the perception is that the new system is an effort to rate more teachers as effective or needs improvement”
- “The state has intentionally made it difficult for teachers to receive ratings of excellent. The belief is that proficient is what people should aspire to and that

excellent should be rare. It creates a demoralizing dynamic for excellent teachers.”

- “The unintended impact is an atmosphere of mistrust and negative feelings on the part of teachers who have always received the highest level rating under the three tier system and now are not.”
- “Teachers have become quite sensitive to the rating system and its impact on their careers and finances. Teachers who view themselves as highly effective, but are not evaluated at that level become frustrated and demoralized.”



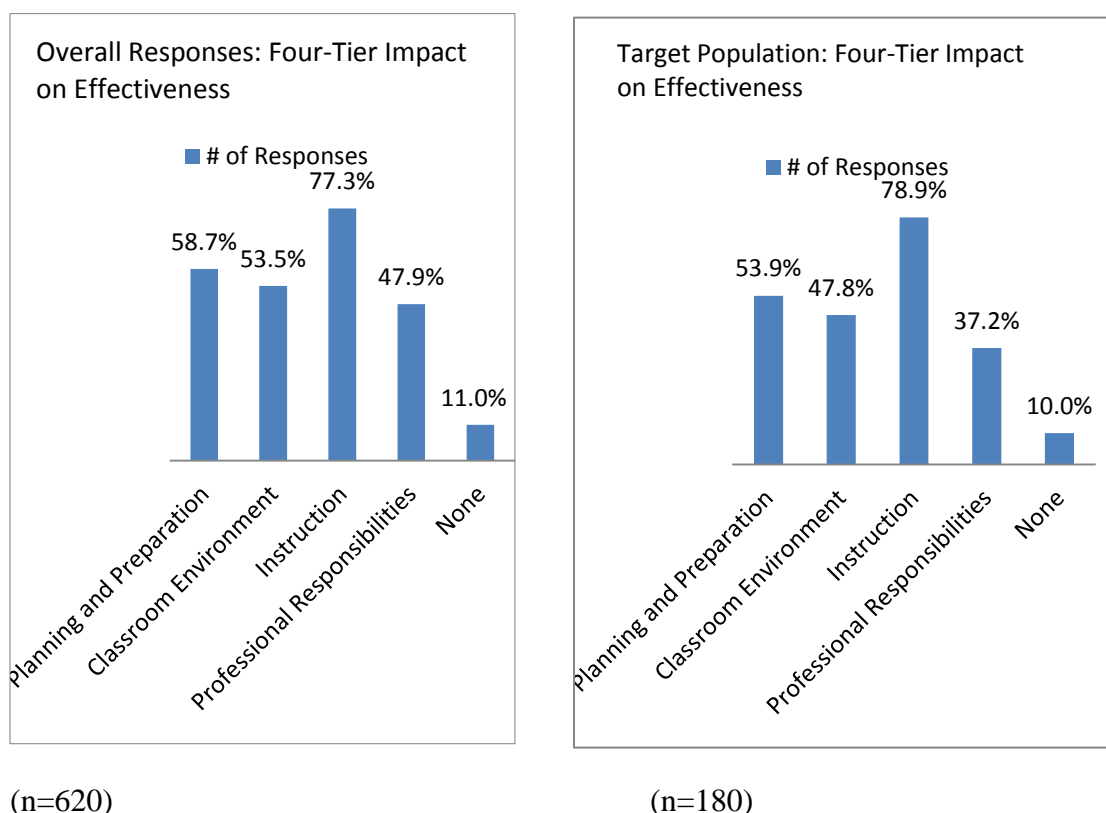
(n=154)

*Figure 17. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Recognition*

Compared to the intended impacts, these unintended impacts were further distributed in regards to their themes. Following teacher morale some of the top themes were interference with teacher growth, teacher competition and comparisons, misconceptions and misunderstandings, too time intensive, teacher fear, poor teacher/administrator relationships, teacher stress, teacher frustration, teacher resentment, tougher criteria, inflation of scores, and promoting teacher growth. The principals were permitted to select multiple categories which were why the total responses exceed 100%.

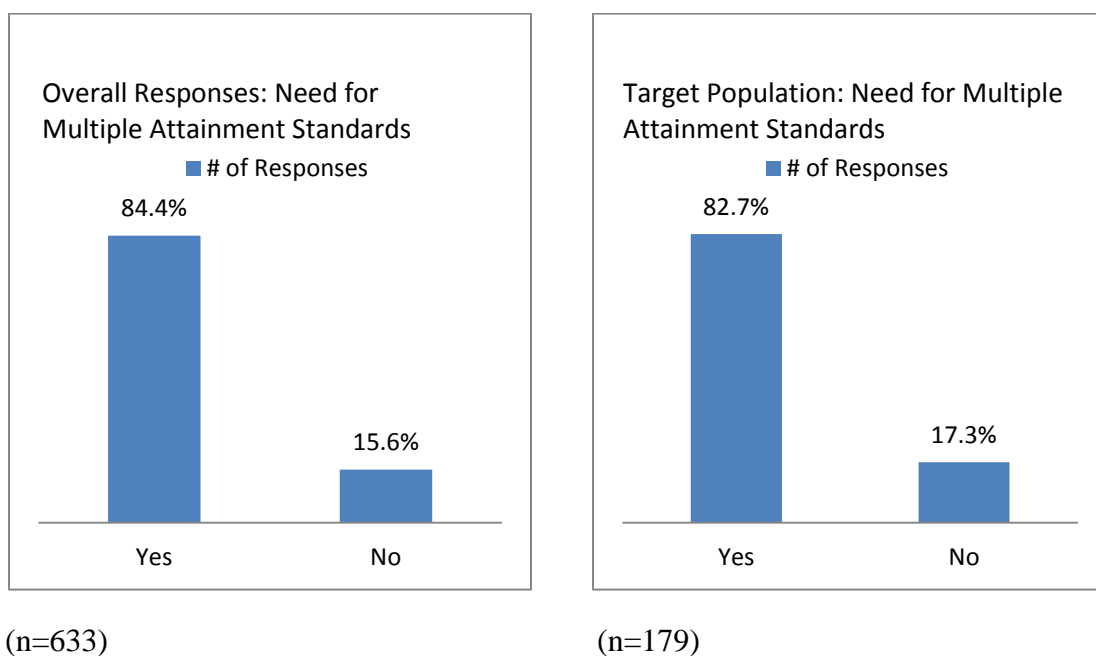
### **Teacher Effectiveness**

The focus of the instrument next shifted to principal perceptions regarding how the expanded four-tier teacher rating system could more effectively identify teacher effectiveness compared to their prior teacher rating system in the domains of planning and preparation, classroom environment, instruction, and professional responsibilities. In Figure 18, similar to the previous question on recognizing teacher excellence, the highest area of impact was teacher instruction (n=479, 77.3%) which was followed by planning and preparation (n=364, 58.7%), classroom environment (n=332, 53.5%), and professional responsibilities (n=297, 47.9%). Some principals felt that none of the domains were able to more effectively identify teacher effectiveness (n=68, 11.0%). The principals were permitted to select multiple categories which were why the total responses exceed 100%. The target population identified instruction as the domain that most effectively identified teacher effectiveness (n=142, 78.9%). Professional responsibilities showed the largest decrease between the overall responses and the target population with a 10.9% decrease (n=67, 37.2%) (see Figure 18).



*Figure 18. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Recognize Teacher Effectiveness*

Principals were asked if it was necessary to have a system with more than one tier to recognize teacher attainment of standards (i.e., “Proficient” and “Excellent”) which is displayed in Figure 19. The vast majority of the overall responses indicated “yes” (n=534) compared to “no” (n=99). The target population saw a similar trend that fluctuated by less than 2% from the overall responses between “yes” and “no” (see Figure 19).



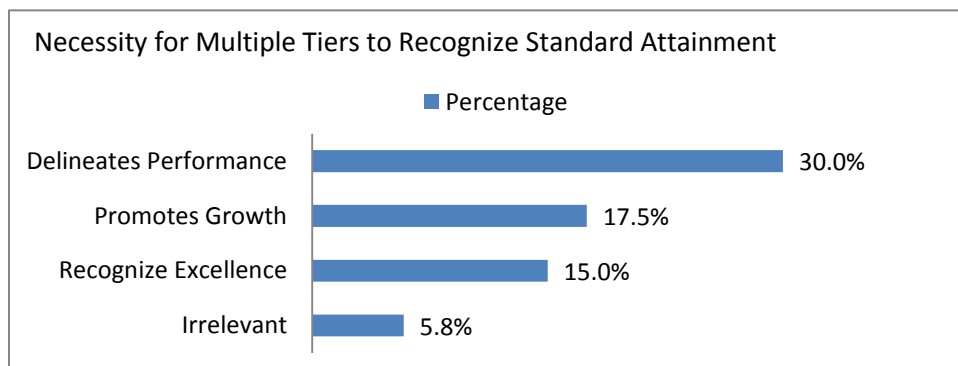
*Figure 19. Necessity for a System with More Than One Tier to Recognize Teacher*

When principals were asked to explain their answers regarding the necessity for more than one tier to recognize teacher attainment of standards there were four themes that emerged. As Figure 20 illustrates, the most frequent response was that the multiple tiers delineated performance (30.0%) as detailed in the following examples:

- “I believe that the levels serve as a scale in the same way we strive to create deeper understanding in students. A tiered system serves a "map" for learning and developing.”
- “Teachers are learners just like students and attainment of standards goes along a continuum so teachers should have the opportunity to show where they are in their learning with all the standards.”

- “I think that different levels reflect the teacher's mastery of those standards and can differentiate the efficacy of their behaviors in affecting student achievement.”
- “Assessment systems should allow for a proficient rating while acknowledging performance that is above the proficient level.”

This was followed by the promotion of teacher growth (17.5%), recognizing excellence (15.0%), and that the multiple tiers were irrelevant (5.8%).



(n=120)

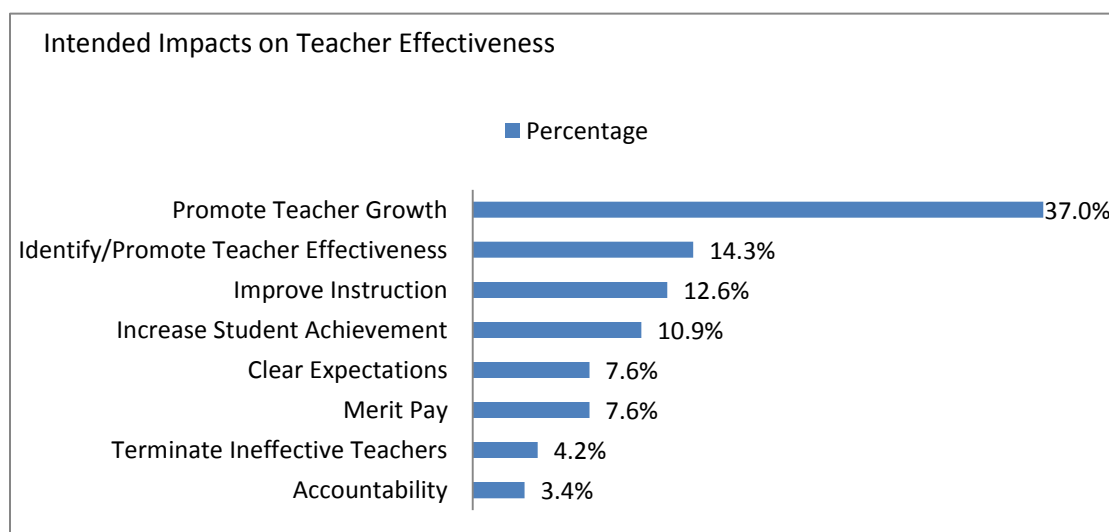
*Figure 20.* Necessity to Have a System with More than One Tier to Recognize Teacher Attainment of Standards (i.e., “Proficient” and “Excellent”)

The principals were next asked to provide a written response regarding what they believed were the intended impacts of the four-tier rating system in relation to teacher effectiveness in their state (see Figure 21). After coding the responses from the target population the highest frequency of responses focused on the promotion of teacher growth (37.0%) as detailed in the following examples:

- “It has caused teachers to examine their planning, lesson delivery and student assessment practices.”

- “Teachers are taking their professional development more seriously. They are actively seeking out ways to better themselves and to better each other. This is not just because of the system is four tier. It is because of the rubric developed to identify those tiers.”
- “To clearly identify what quality instruction looks like for all teachers and then help them attain those skills. Being consistent with these across the board will theoretically give all students the same high level of instruction.”
- “To help teachers better understand their strengths and challenges so that they can improve their professional practice.”

This theme was followed by identifying and promoting teacher effectiveness, improving instruction, increasing student achievement, providing clear expectations, merit pay, terminating ineffective teachers, and greater accountability.



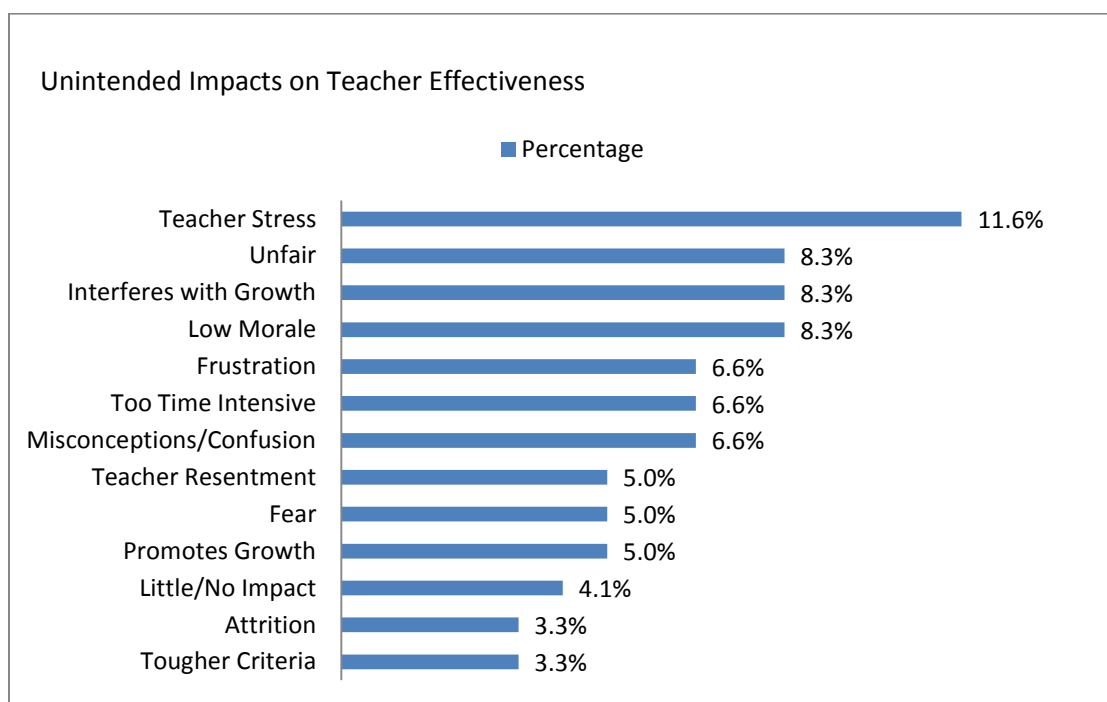
(n=119)

*Figure 21.* The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Effectiveness

The principals were asked to provide extended responses regarding what they believed were the unintended impacts of the four-tier rating system in relation to teacher effectiveness in their state (see Figure 22). This response coding from the target population revealed the most frequent response was related to teacher stress (11.6%) as detailed in the following examples:

- “Funding and time are needed in order to provide professional development for teachers to hone their skills. Using the system to evaluate teachers before providing the expectations, intended outcomes, and training has been a disservice to teachers. It has created a lot of stress and anxiety that would have been alleviated if the plan was implemented in the proper sequence.”
- “The profession is becoming less desirable to many considering entering or staying in teaching careers. The pressure to improve quickly and meet higher levels of performance is great. Pay is already too low and the model makes it harder to move up on the pay scale.”
- “Teacher stress due to accountability from the tiers and rubrics. Teachers once rated proficient may not be based on the rubrics.”

The other themes that were discovered from this question were widely distributed as teacher stress was followed by unfair evaluations, interference with growth, low morale, teacher frustration, too time intensive, misconceptions and confusion, promoting teacher growth, teacher resentment, teacher fear, little to no impact, teacher attrition, and tougher criteria.



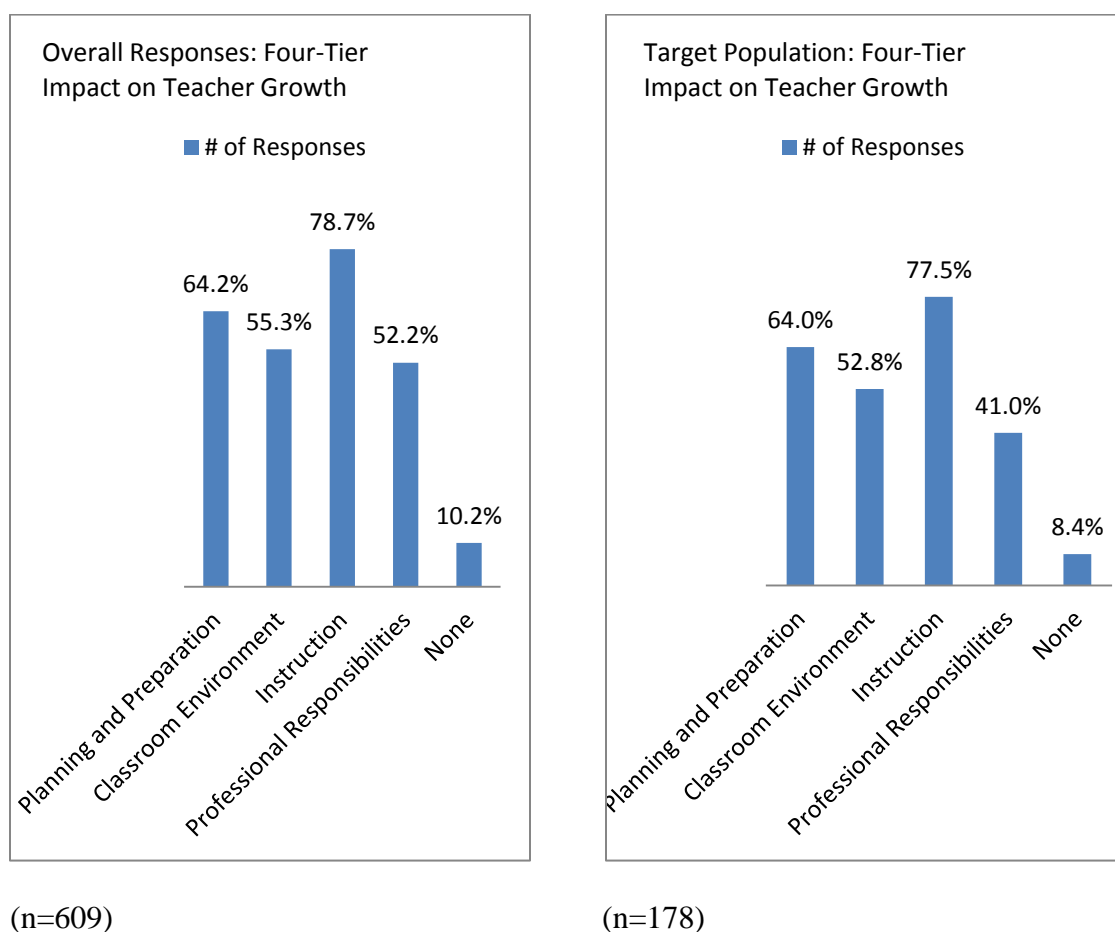
(n=121)

*Figure 22. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Effectiveness*

### Teacher Growth

Principals were next asked their perceptions regarding how the expanded four-tier teacher rating system promoted teacher growth compared to their prior teacher rating system in the domains of planning and preparation, classroom environment, instruction, and professional responsibilities. Similar to the previous question on recognizing teacher excellence and teacher effectiveness, the highest area of impact was teacher instruction (n=479, 78.7%) which was followed by planning and preparation (n=391, 64.2%), classroom environment (n=318, 55.3%), and professional responsibilities (n=318, 52.2%) as exhibited in Figure 23. Again, some principals felt that none of the domains were able to more effectively promote teacher growth in the expanded four-tier system (n=62,

10.2%). Once again, the target population was lower for every subgroup as it was with recognizing teacher excellence. In Figure 23, the target population still identified instruction as the domain that most effectively promoted teacher growth (n=138, 77.5%). Once again, professional responsibilities showed the largest decrease between the overall responses and the target population with a 10.7% decrease (n=73, 41.0%).



*Figure 23. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Promote Teacher Growth*

In regards to how the principals believed the performance ratings they assigned a given teacher impacted teacher growth (see Figure 24), there was a clear majority in the responses. The principals said the performance rating given “usually promotes” teacher

growth in both the overall responses (n=470, 77.7%) and the target population (n=118, 73.3%) (see Figure 24). The number of principals that felt the rating had no impact was the next highest category for the overall responses (13.7%) and the target population (17.4%). Even fewer principals felt the performance rating distracted from teacher growth for the overall responses (4.9%) and the target population (5.6%).

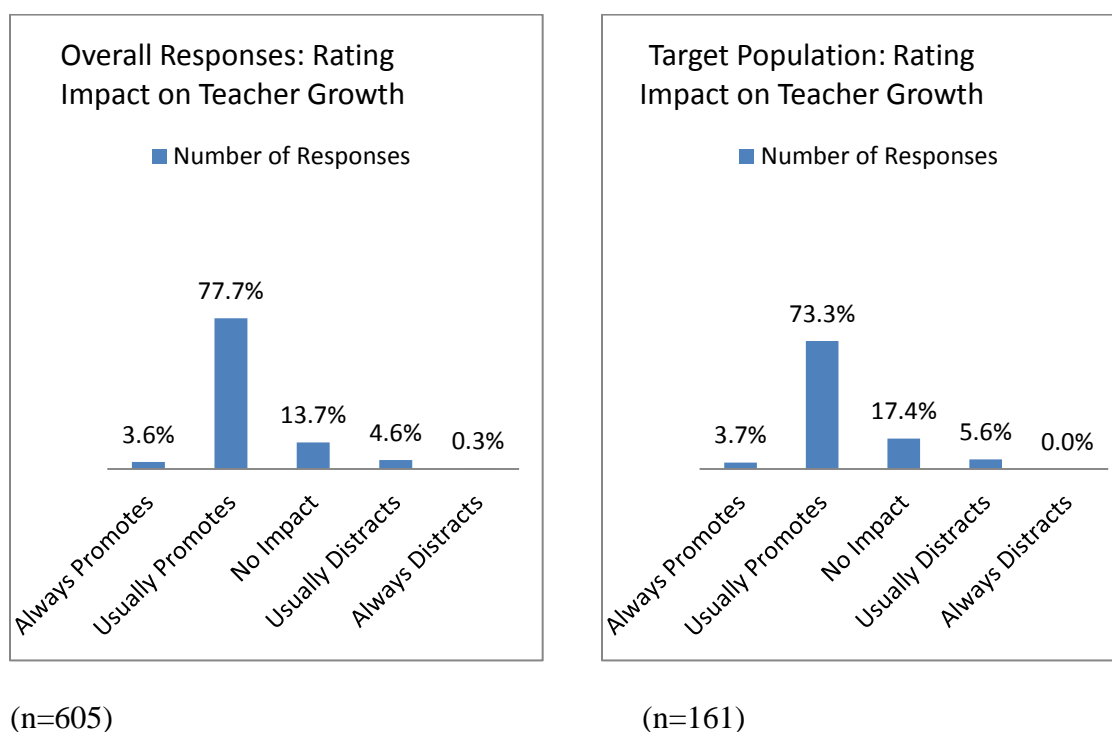
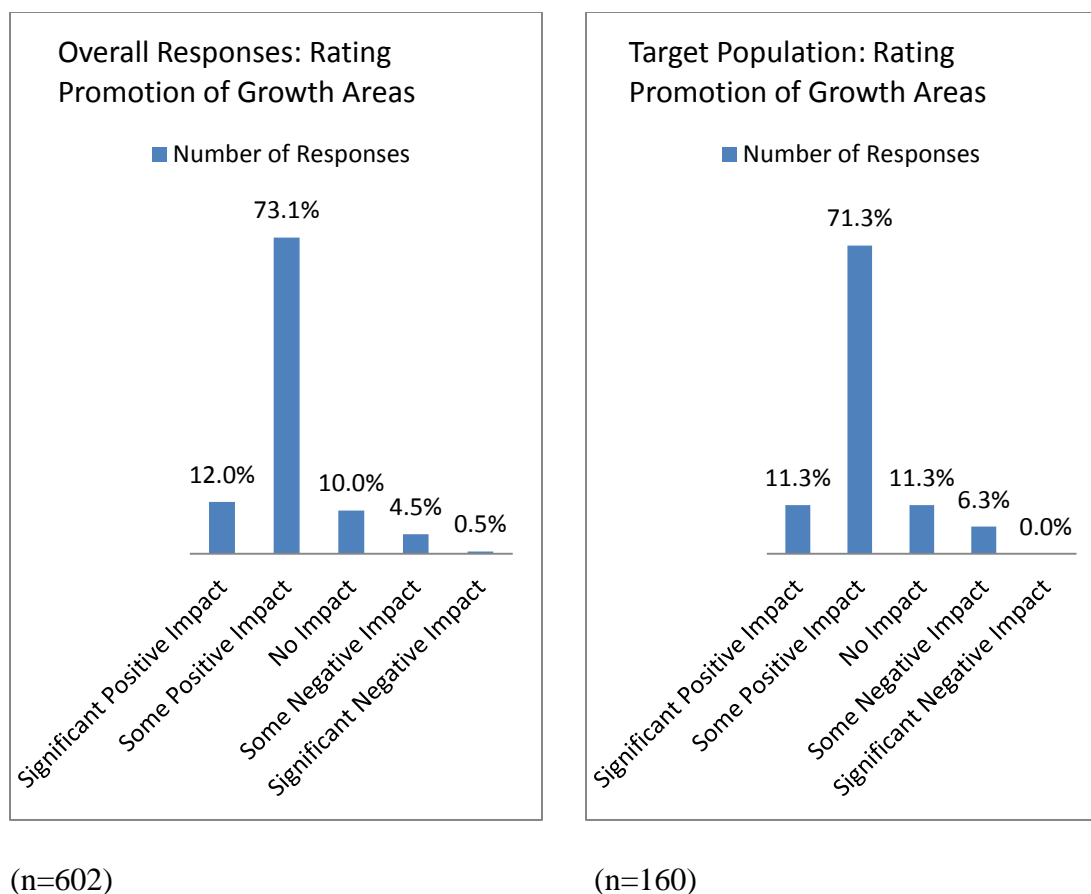


Figure 24. Impact of the Performance Rating on Teacher Growth

Similar trends were found when the principals were asked if they felt the performance ratings promoted the identified areas for growth in the evaluation as displayed in Figure 25. The principals said the performance rating given had “some positive impact” on promoting identified areas for teacher growth in both the overall responses (n=440, 73.1%) and the target population (n=114, 71.3%) (see Figure 25). The number of principals that felt the rating had no impact on promoting identified areas for

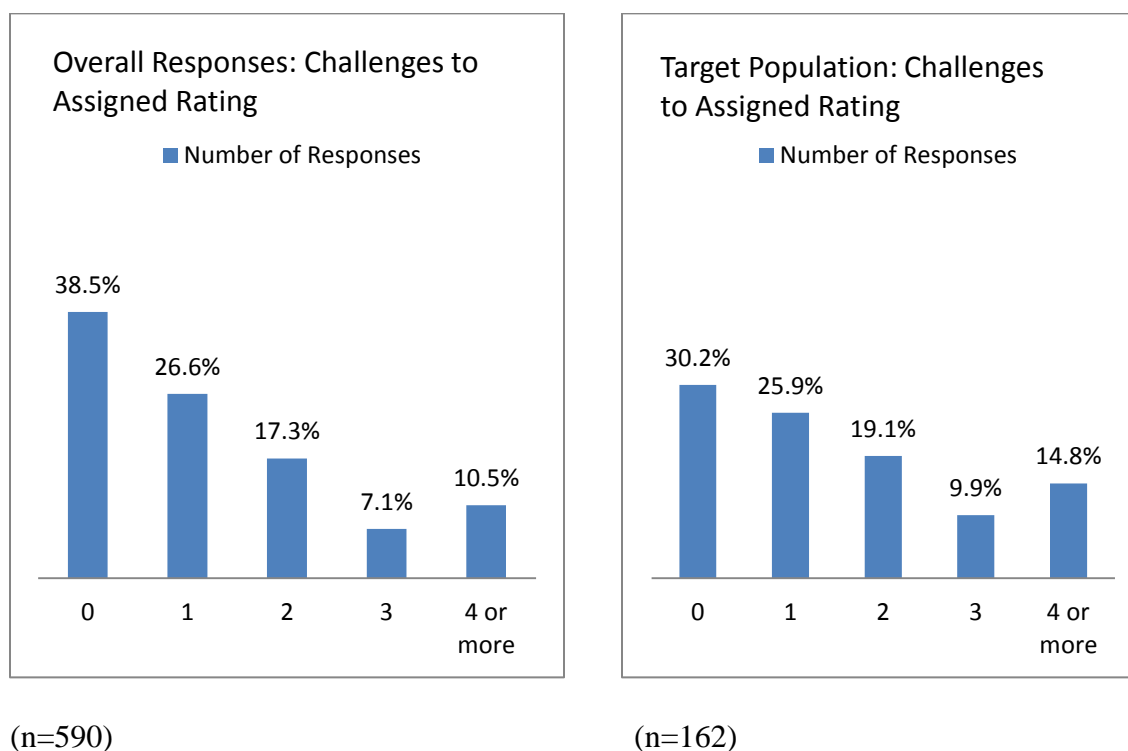
growth was similar for the overall responses (10.0%) and the target population (11.3%). Even fewer principals felt the performance rating had a negative impact on the identified areas for teacher growth for the overall responses (4.5%) and the target population (6.3%).



*Figure 25. Performance Rating Promotion of Identified Areas for Growth in the Evaluation*

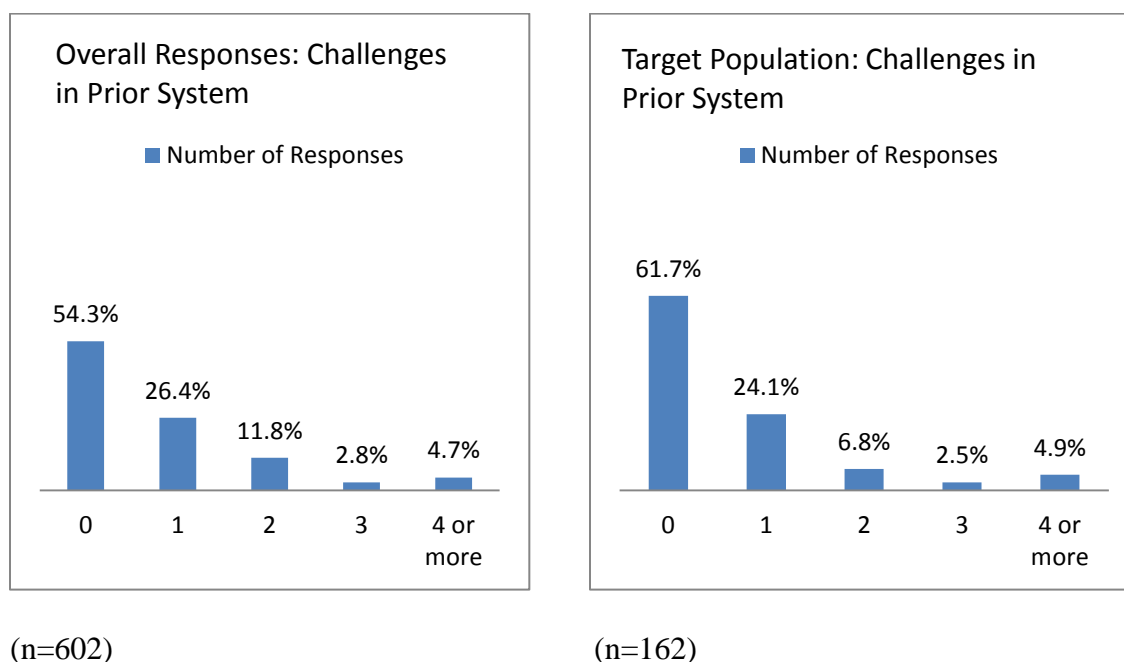
The principals provided a wide range of responses in regards to how many times per year teachers have challenged or protested the performance ratings they have assigned teachers under the four-tier rating system (see Figure 26). The majority of principals had 0-1 challenges to the assigned rating for both the overall responses

(65.1%) and the target population (56.1%) (see Figure 26). Principals reporting more than one challenge to the assigned teacher rating was higher in the target population for all subgroups in comparison to the overall responses.



*Figure 26. Frequency of Teacher Challenges or Protests of Assigned Performance Rating Under the Four-Tier System*

In comparison, the principals were asked how many times per year they encountered a challenge or protest to the performance rating they assigned under their previous evaluation system as displayed in Figure 27. In comparison to the four-tier system (see Figure 26), there were far less challenges to the assigned performance rating in the principals' previous evaluation system. The majority of principals had 0-1 challenges to the assigned rating at a significantly higher rate for both the overall responses (80.7%) and the target population (85.8%).



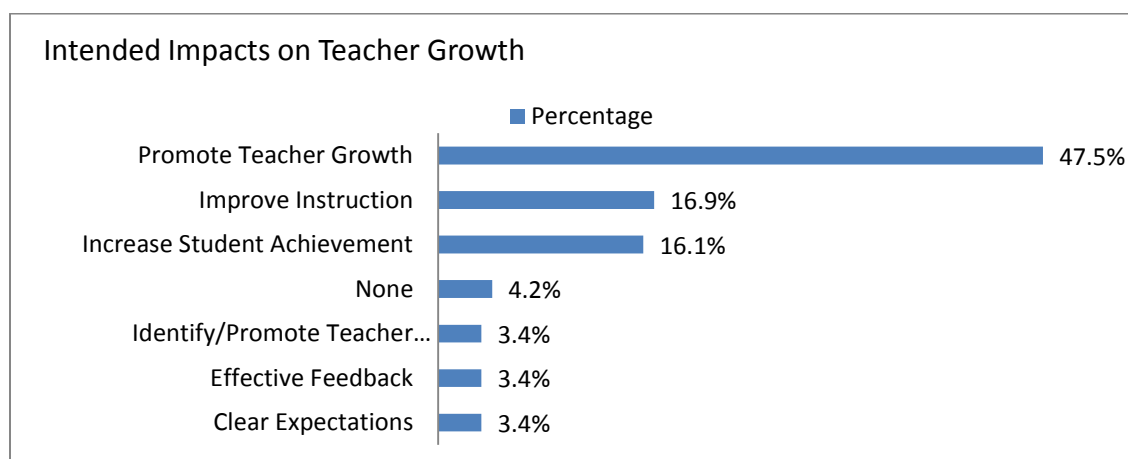
*Figure 27. Frequency of Teacher Challenges or Protests of Assigned Performance Rating Under the Principals' Previous Evaluation System*

The principals were next asked to provide extended responses regarding what they believed were the intended impacts of the four-tier rating system in relation to teacher growth in their state (see Figure 28). Nearly half of the responses focused on the promotion of teacher growth (47.5%) as detailed in the following examples:

- “To ensure that teachers continue to grow in their profession and apply new strategies research based strategies into their daily instruction.”
- “The system requires teachers to examine their practice more regularly and more deeply.”
- “To clearly define areas that need improvement and provide concrete suggestions for improvement.”

- “Our developmental practice is designed to promote meta-cognition allowing teachers to take responsibility for their own learning and growth.”
- “The implementation of the system has resulted in (1) an increase in collaboration between administrators and teachers as well as between teachers; (2) more effectively addressing the needs of students; and (3) a focus on professional growth and development.”

This theme was followed by improving instruction, increasing student achievement, no intended impacts on teacher growth, identifying and promoting teacher effectiveness, providing effective feedback, and clear expectations.



(n=118)

*Figure 28.* The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Growth

The principals were asked to provide extended responses regarding what they believed were the unintended impacts of the four-tier rating system in relation to teacher growth in their state (see Figure 29). This response coding from the target population revealed the most frequent response was related to interference with growth (15.9%)

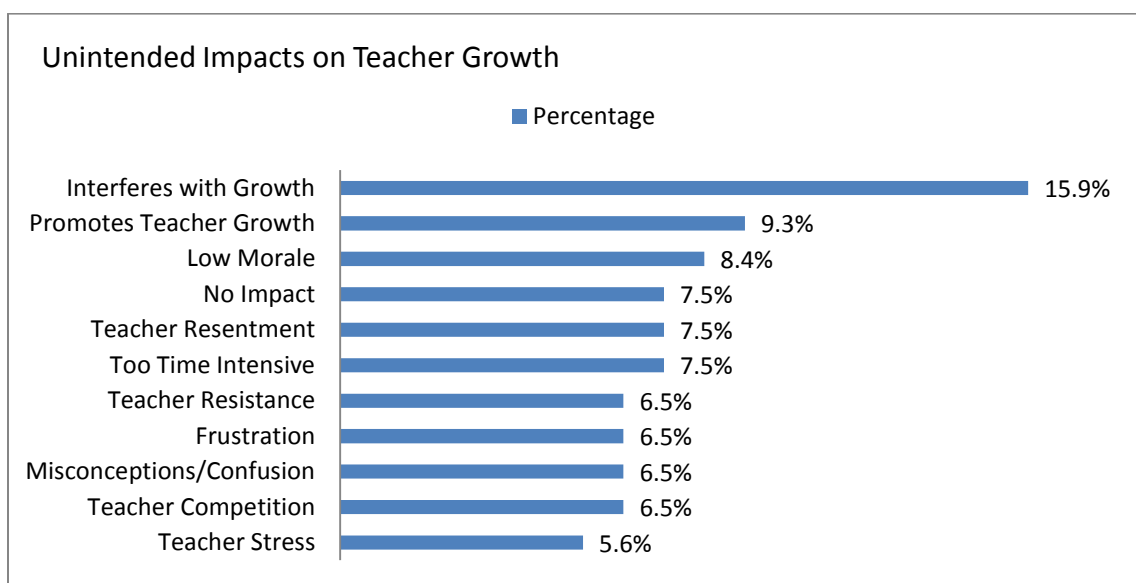
which was followed in frequency by the opposite concept of promoting teacher growth (9.3%). Comments related to the interference with teacher growth are detailed in the following examples:

- “Some teachers just aren't ready for change and so they get caught up on the rating instead of how they can grow as a professional and educator.”
- “Stress over system undermines growth in some teachers.”
- “They want to do what they have to do to earn the highest rating, and they are not thinking about growth.”
- “Paperwork may get in the way of time for growth opportunities.”
- “Under performing teachers will use the union to avoid difficult and honest conversation regarding areas in need of improvement.”

Comments related to the promotion of teacher growth as an unintended consequence are detailed in the following examples:

- “Teachers can't hide behind something that they did years ago - they need to be constantly making improvements. This is a good thing.”
- “More involvement in teacher led learning communities.”
- “Professional growth and learning with teacher collaboration.”

The next unintended impact theme was low morale which was followed in frequency no unintended impact, teacher resentment, too time intensive, teacher resistance, teacher frustration, misconceptions and confusion, teacher competition, and teacher stress.



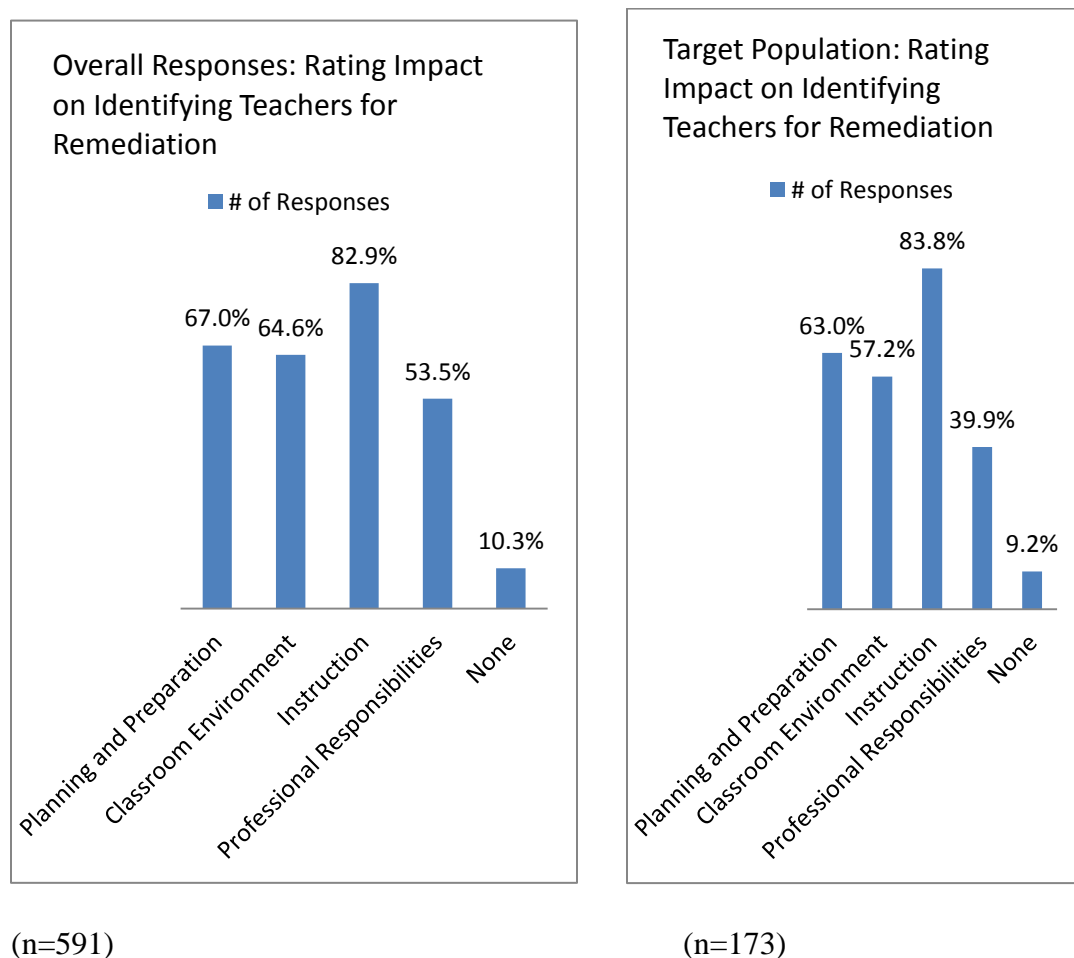
(n=107)

*Figure 29. The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Growth*

### **Teacher Remediation**

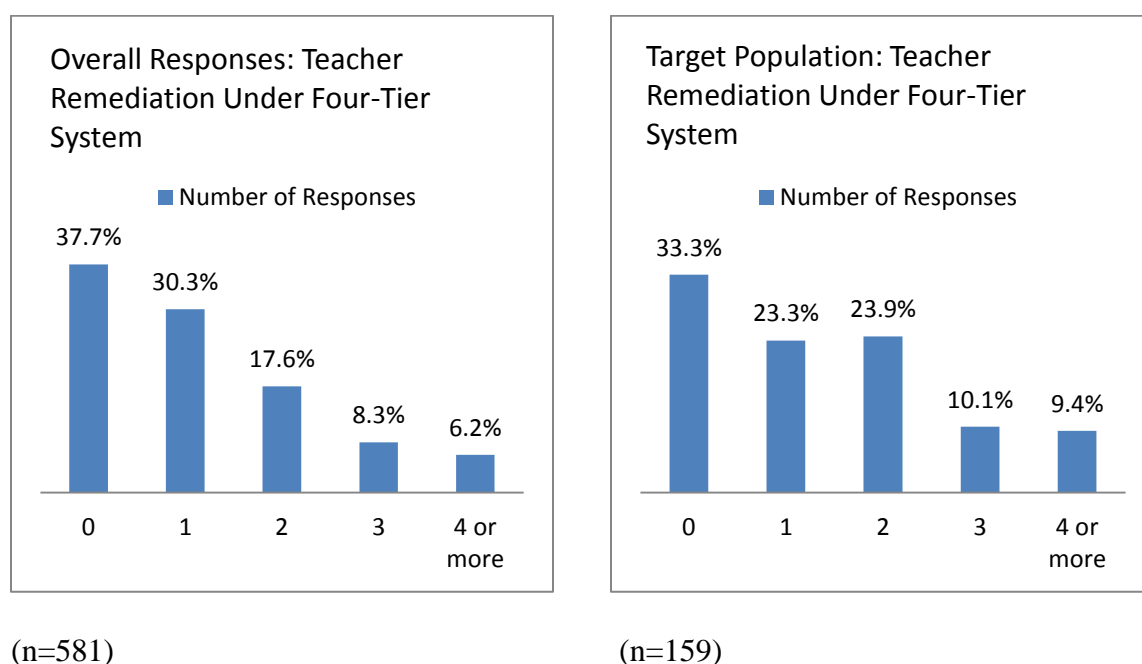
Principals were asked their perceptions regarding how the expanded four-tier teacher rating system promoted teacher remediation compared to their prior teacher rating system in the domains of planning and preparation, classroom environment, instruction, and professional responsibilities. Once again, similar to the previous question on recognizing teacher excellence and teacher growth the highest area of impact was teacher instruction (n=490, 82.9%) which was followed by planning and preparation (n=396, 67.0%), classroom environment (n=382, 64.6%), and professional responsibilities (n=316, 53.5%) as exhibited in Figure 30. Some principals felt that none of the domains were able to more effectively promote teacher remediation (n=61, 10.3%). The target population was lower for planning and preparation (63.0%), classroom environment

(57.2%), and professional responsibilities (39.9%). The target population still identified instruction as the domain that most effectively promoted teacher remediation and it was at a higher rate in comparison to the overall responses (n=145, 83.8%) (see Figure 30). Professional responsibilities once again showed the largest decrease between the overall responses and the target population with a 13.6% decrease (n=69, 39.9%).



*Figure 30. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Identify Teacher for Remediation*

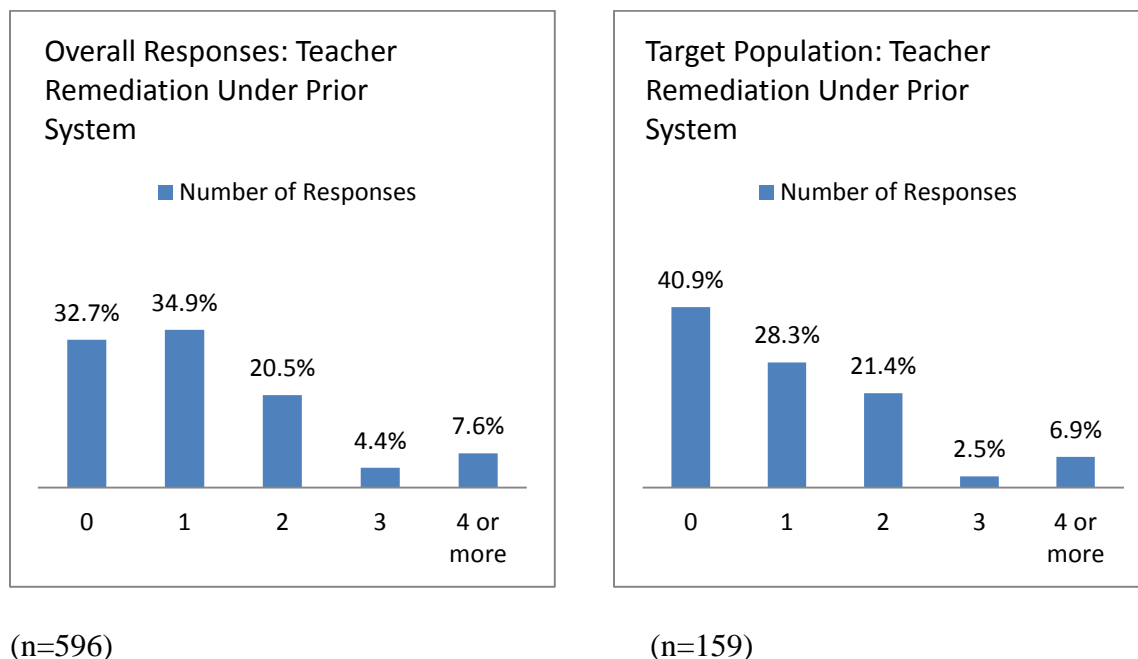
Figure 31 shows that the principals indicated a downward trend in the number of teachers they have put on remediation per year under their current four-tier rating system from zero, on one end of the spectrum, to four or more teachers on the other end. The majority of principals put 0-1 teachers on remediation for both the overall responses (68.0%) and the target population (56.6%). Principals reporting more than one teacher placed on remediation under the current four-tier system were higher in the target population for all subgroups in comparison to the overall responses.



*Figure 31.* Teachers per Year Put on Remediation under the Current Four-Tier Rating System

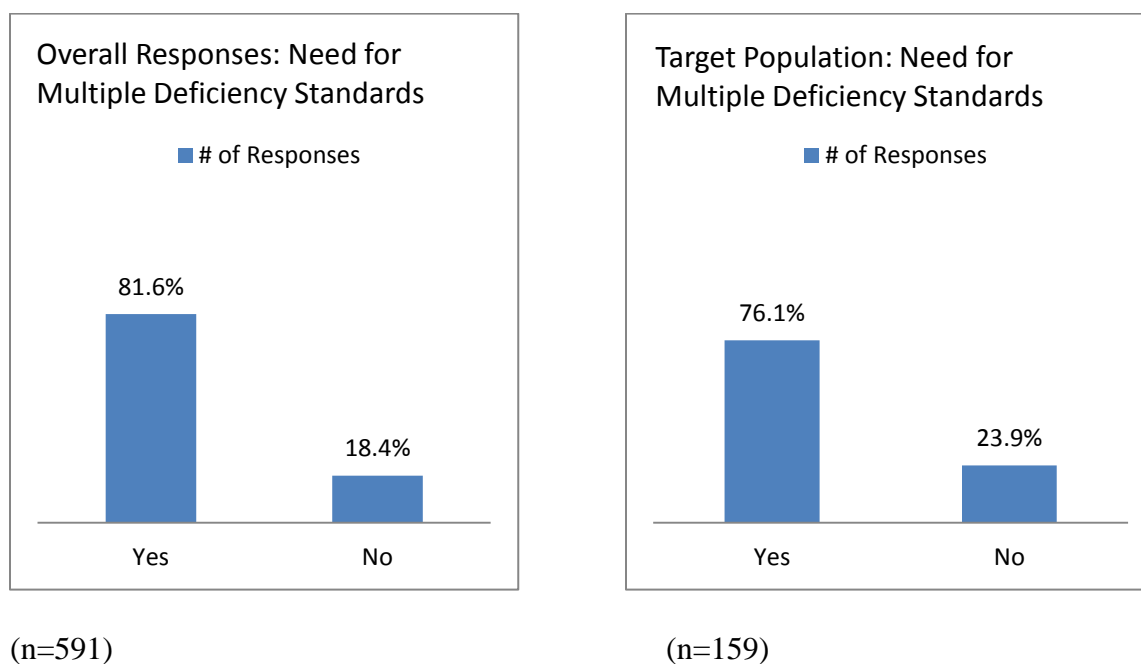
The principals were also asked how many times per year they put teachers on remediation per year under their previous evaluation system as displayed in Figure 32. In comparison to the four-tier system (see Figure 31), there were far less challenges to the assigned performance rating in the principals' previous evaluation system. The majority

of principals had 0-1 teachers placed on remediation per year for both the overall responses (67.6%) and the target population (69.2%).



*Figure 32. Teachers per Year Put on Remediation under the Previous Evaluation System*

Principals were asked if it was necessary to have a system with more than one tier to identify teacher deficiency in meetings standards (i.e., “Needs Improvement” and “Unsatisfactory”) which is displayed in Figure 33. The vast majority of the overall responses indicated “yes” (81.6%, n=482) compared to “no” (18.4%, n=109). The target population responded with a smaller number agreeing to the need for multiple deficiency standards with 5.5% less principals agreeing with the statement (see Figure 33).

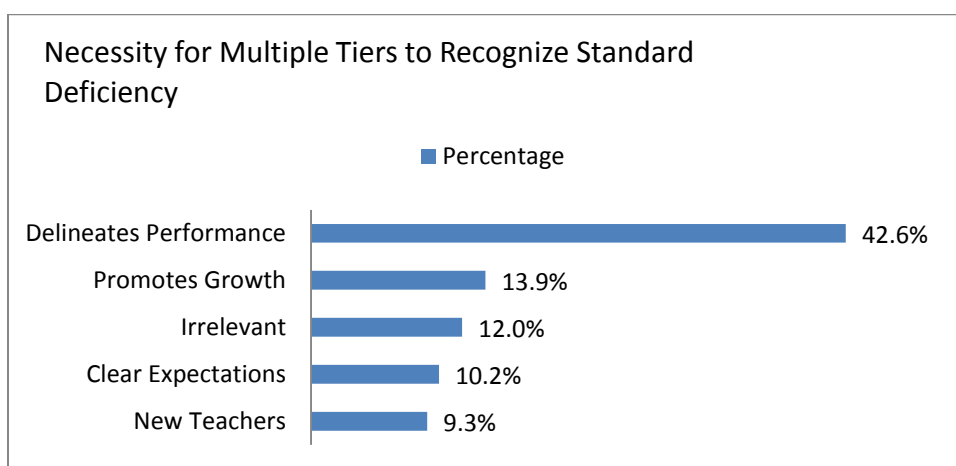


*Figure 33. Necessity for a System with More Than One Tier to Identify Standard Deficiency*

When principals were asked to explain their answers regarding the necessity for more than one tier to recognize teacher deficiency in meeting standards there were five themes that emerged. As Figure 34 illustrates, the most frequent response was that the multiple tiers delineated performance (42.6%) as detailed in the following examples:

- “There is a clear distinction between the teacher that is not making an effort to get better and those that at the very least try to better their teaching skills and instructional delivery.”
- “Our district's system utilizes the terms needs improvement and developing as the tier between unsatisfactory and effective. This distinction has proven useful when evaluating a newer teacher (using the developing term) versus needs improvement for a longer-tenured teacher.”

- “Needs improvement is for teachers who can improve through professional development; unsatisfactory is for teachers who should never be in the classroom, will never meet expectations.”
- “It was possible to identify teacher deficiencies before, it just is more clearly delineated in the four tier system.”
- “To acknowledge the difference between a teacher that needs minimal remediation from a teacher that is going to need significant remediation or possibly dismissal.”



(n=108)

*Figure 34.* Necessity to Have a System with More than One Tier to Recognize Teacher Deficiency in Meeting Standards (i.e., “Needs Improvement” and “Unsatisfactory”)

As the comments illustrate, delineating performance is a predominantly stated argument for multiple tiers to categorize and identify how teachers perform. The second most frequent response, promotion of teacher growth (13.9%), supports the argument that

the multiple tiers also encourages growth and movement within the delineated performance categories as evidenced by the following comments:

- “Teacher performance can fluctuate, and this system enables greater movement of feedback tied to performance on an ongoing basis. As our professional responsibilities, strategies, and curriculum standards change, we too must evolve to expect continuous improvement. To solely expect our student to demonstrate improvement, and not our teachers or administrators is very hypocritical.”
- “Teachers need the opportunity to improve if needed. They need to know where they rate then how to fix it.”
- “A teacher can need improvement without being unsatisfactory.”

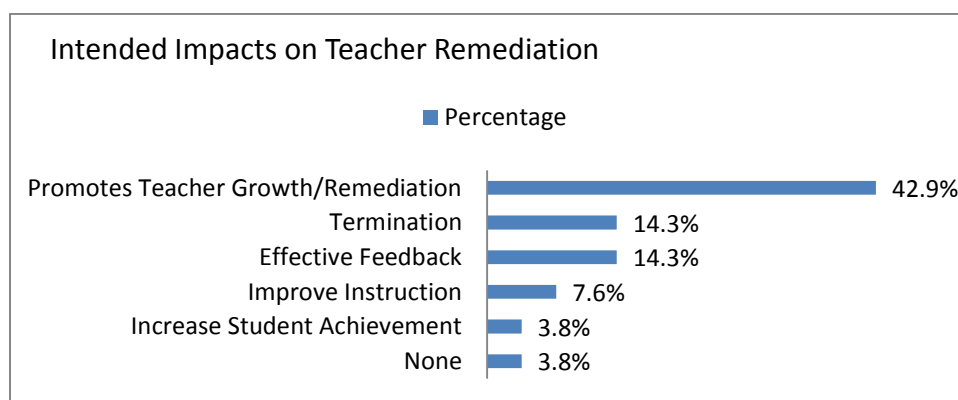
This category was followed by those that believed one or two ratings were irrelevant (12.0%), the existence of clear expectations (10.2%), and the need for multiple tiers for new teachers (9.3%).

The principals were next asked to provide extended responses regarding what they believed were the intended impacts of the four-tier rating system in relation to teacher remediation in their state (see Figure 35). The largest concentration of responses again focused on the promotion of teacher growth (42.9%) as detailed in the following examples:

- “To allow teachers the opportunity to improve their instruction.”
- “To enable new teachers time to grow without penalty, veteran personnel to adapt to changes, and ultimately everyone focus on continuous improvement.”

- “I think the intention is to identify and then provide support in a variety of ways so that the teacher has the opportunity to grow and improve.”
- “The intent is to make it easier to identify teachers that need help and support, give them that support, and remove them from the classroom in a timely manner if teaching, with supports, does not improve.”

This theme was followed by terminating ineffective teachers, providing effective feedback, improving instruction, increasing student achievement, and those that felt there were no intended impacts in relation to teacher remediation.



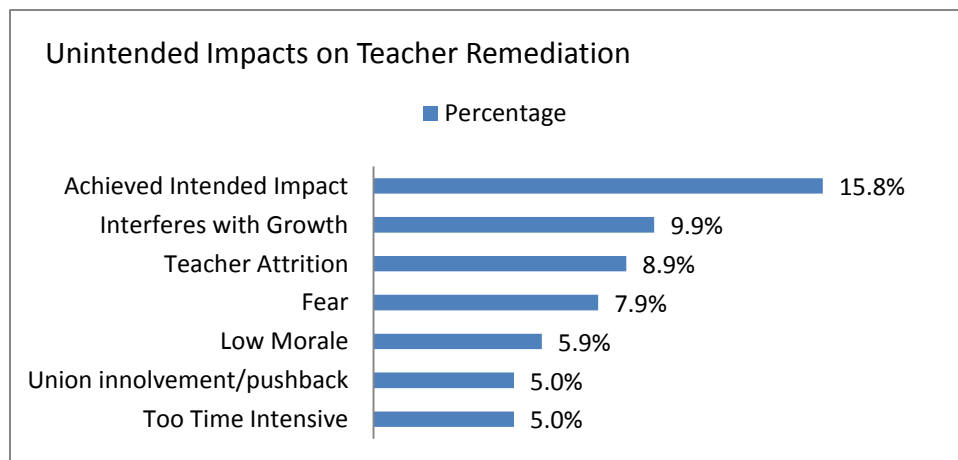
(n=104)

*Figure 35.* The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Remediation

The principals were asked to provide extended responses regarding what they believed were the unintended impacts of the four-tier rating system in relation to teacher remediation in their state (see Figure 36). The response coding from the target population was widely distributed but revealed the most frequent response was that there were no unintended consequences related to teacher remediation (15.8%) as detailed in the following examples:

- “I haven't really seen unintended impacts related to teacher remediation. It has been an effective tool for me to bring people's attention to where they need to grow.”
- “No difference in the % of teachers evaluated with a low rating.”
- “The demands on the principal to carry out this system procedurally will limit who to focus on instead promoting growth across the core of instructional staff. In the end the use of and outcomes from the system will mimic the old system.”

The themes that followed in frequency were that it interferes with teacher growth, teacher attrition, teacher fear, low morale, union involvement and pushback, and that the process is too time intensive.

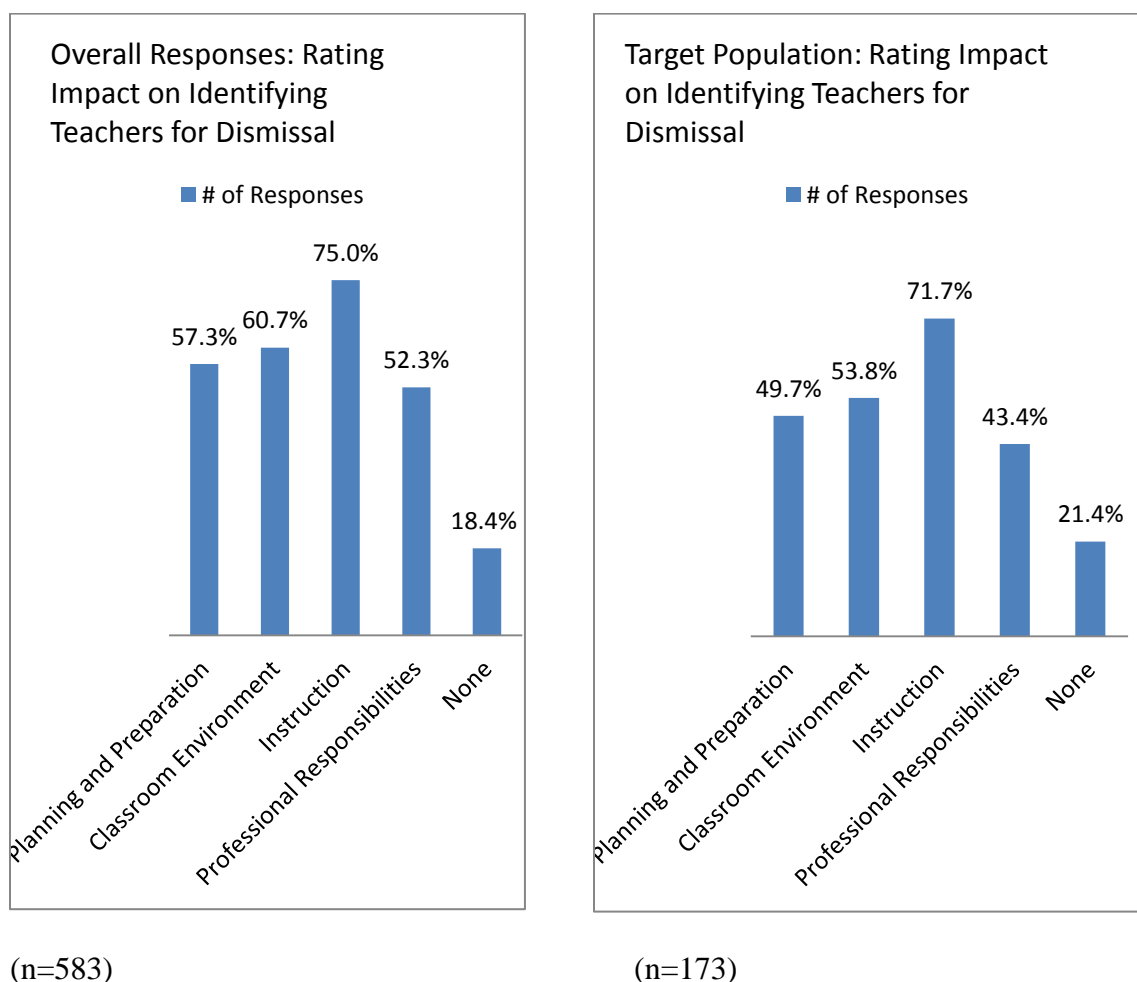


(n=101)

*Figure 36.* The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Remediation

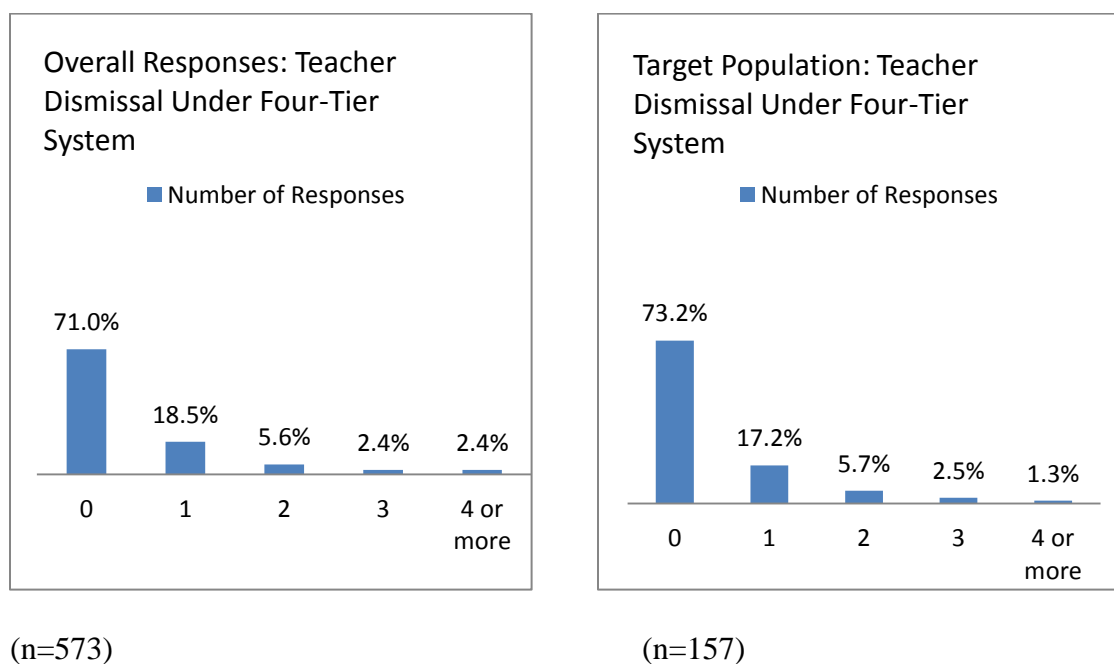
### **Teacher Dismissal**

Lastly, principals were asked about their perceptions regarding how the expanded four-tier teacher rating system could more effectively identify teachers for dismissal compared to their prior teacher rating system in the domains of planning and preparation, classroom environment, instruction, and professional responsibilities. Similar to the previous questions on recognizing teacher excellence, teacher effectiveness, teacher growth, and teacher remediation, the highest area of impact was teacher instruction (n=437, 75.0%) which was followed by classroom environment (n=354, 60.7%), planning and preparation (n=334, 57.3%), and professional responsibilities (n=305, 52.3%) as exhibited in Figure 37. Some principals felt that none of the domains were able to more effectively promote teacher growth (n=107, 18.4%). The target population was lower for planning and preparation (49.7%), classroom environment (53.8%) and professional responsibilities (43.4%) (see Figure 37). The target population still identified instruction as the domain that most effectively identified teachers for dismissal and it was at a lower rate in comparison to the overall responses (71.7%). Professional responsibilities once again showed the largest decrease between the overall responses and the target population with an 8.9% decrease (43.4%).



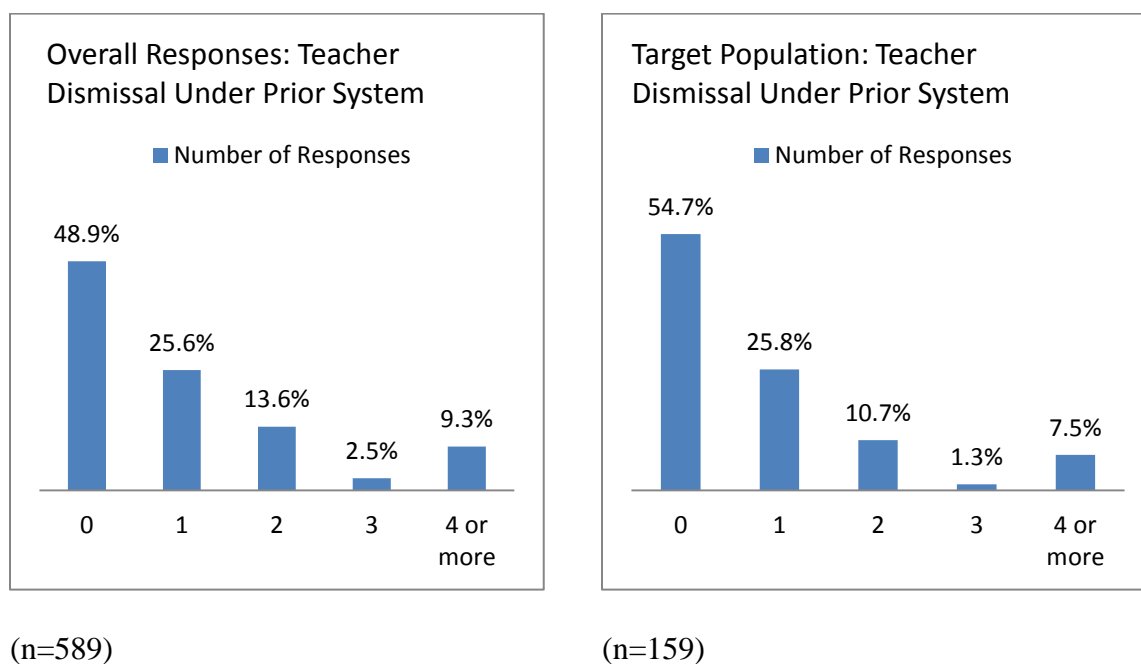
*Figure 37. Domains in Which Principals Believe Their Expanded Four-Tier System Can More Effectively Identify Teacher for Dismissal*

Figure 38 shows that the responses of principals indicated a downward trend in the number of teachers they have dismissed per year under their current four-tier rating system from “zero” to “four or more” teachers. The majority of principals put 0-1 teachers on remediation for both the overall responses (71.0%) and the target population (73.2%). Principals reporting more than one teacher dismissed under the current four-tier system showed little variation in the target population for all subgroups in comparison to the overall responses.



*Figure 38. Teachers per Year Terminated under the Current Four-Tier Rating System*

The principals were also asked how many times per year they dismissed teachers under their previous evaluation system, as displayed in Figure 39. In comparison to the four-tier system (see Figure 38), there were significantly more dismissals under the prior rating system for both the target population and the overall responses according to the principals surveyed. The number of principals that reported they dismissed zero teachers per year under their previous model was lower for the overall population (n=288, 48.9%) and the target population (n=87, 54.7%) in comparison to the same question with their current model. Conversely, the numbers were higher for all but one other subcategory from one to four or more teachers dismissed per year in the previous model compared to the current model.



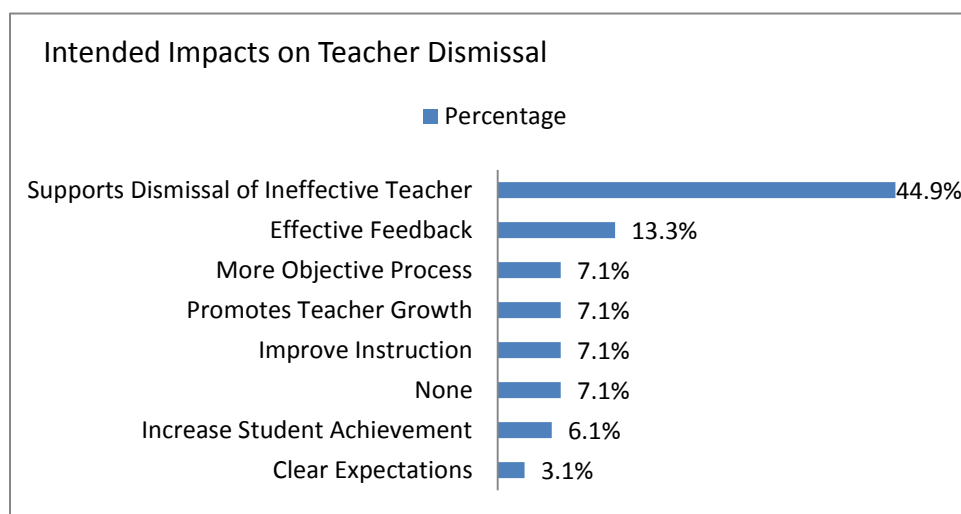
*Figure 39. Teachers per Year Dismissed under the Previous Evaluation System*

The principals were next asked to provide extended responses regarding what they believed were the intended impacts of the four-tier rating system in relation to teacher dismissal in their state (see Figure 40). The largest concentration of responses related to the system supporting the dismissal of ineffective teachers (44.9%) as detailed in the following examples:

- “The intended impact was to create a dismissal system based on performance rather than provide protections through tenure.”
- “If a teacher doesn't improve with support over a specified time, they would be recommended for dismissal.”
- “The intended impact is to provide quality teachers in each classroom and remove those who aren't quality teachers.”

- “It should allow greater opportunities for remediation and professional growth which reduce the instance of termination. However, if attempts to remediate unsatisfactory areas failed, then termination should be an easier process.”
- “The intended impact was for chronically underperforming teachers be terminated after a reasonable period of time attempting remediation of poor practice.”

This theme was followed by providing effective feedback, a more objective process, improving instruction, no intended impact, increasing student achievement, and clear expectations.



(n=98)

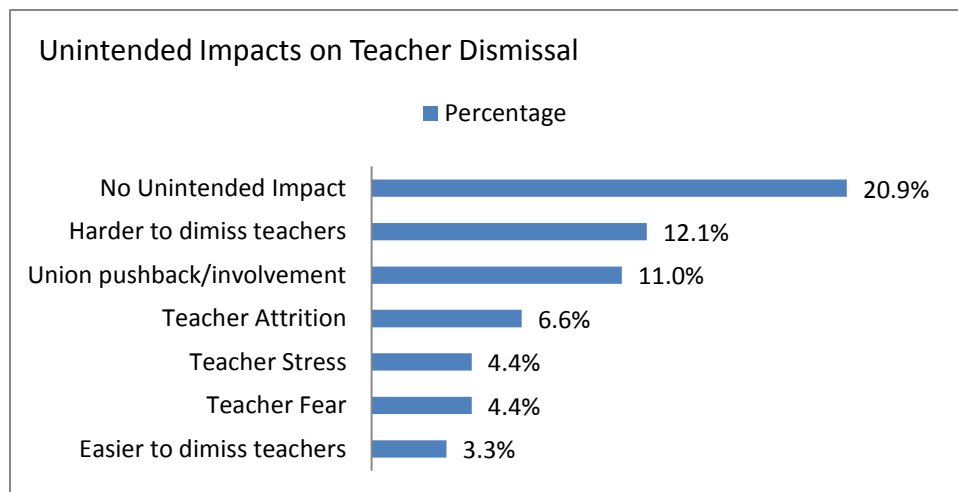
*Figure 40.* The Intended Impacts of the Four-Tier Rating System in Relation to Teacher Dismissal

The principals were asked to provide extended responses regarding what they believed were the unintended impacts of the four-tier rating system in relation to teacher dismissal in their state (see Figure 41). These responses indicated that the most common

theme was no unintended consequences related to teacher dismissal (20.9%) as detailed in the following examples:

- “There do not appear to be any unintended impacts at this time. Both teachers and evaluators appear to understand that the intended impact is the actual impact.”
- “None”
- “Union contract language still prevails and the four tier system does nothing to assist with teacher dismissal.”

The themes that followed in frequency were that it was harder to dismiss teachers, union pushback and involvement, teacher attrition, teacher stress, teacher fear, and that it was easier to dismiss teachers.



(n=91)

*Figure 41.* The Unintended Impacts of the Four-Tier Rating System in Relation to Teacher Dismissal

**Additional Comments**

The last question in the survey welcomed the respondents to further explain any answers they provided in the survey. Many principals shared insights regarding their experiences to this point with the four-tier system in their respective state and how they felt it impacted students, teachers, and administrators. In analyzing the target population specifically there was no significant themes that emerged in these responses. Here are some of the examples that were shared:

- “The tool we use has the clear potential to be "game changing", but the tiered system that it is tied to has completely undermined that potential.”
- “I believe educators want to be successful and have high student achievement, we have to continue to strive to improve the learning that is happening in our buildings, evaluation ratings are just a subjective process that in my opinion do little to improve the learning that happens. As a principal, it is my responsibility to strive to improve our craft everyday.”
- “Teachers cannot hide within their four walls any longer. They need to be transparent and collaborate with each other. Student achievement is our goal!”
- “Our 4 tier system requires much more time from administrators. I find the conversations quite useful. Too much pressure is put on administrators to find teachers in need of improvement. Some of the unintended consequences are a result of using the system to determine teacher pay. If teacher pay was at an acceptable level in the first place, the consequence would not have such a negative impact.”

- “In my 25 years of educational experience I have yet to see anything treat teachers so unfairly, and reduce the positive impact good teachers will have on their students. Law makers should stay away from education.”
- “I have seen some positive outcomes with the new system. The progress has been slow and the time to evaluate teachers is very time consuming.”

### **Conceptual Framework**

The principal perceptions are displayed below using Danielson’s four domains as a conceptual framework to understand these data for each of the five areas being studied (recognition, effectiveness, growth, remediation, and dismissal) (see Tables 11 and 12). In surveying Massachusetts and Florida the four domains of: 1) planning and preparation, 2) classroom environment, 3) instruction, and 4) professional responsibilities were measured in relation to teacher recognition, effectiveness, growth, remediation, and dismissal using the four-tiered performance rating system.

As was previously discussed in this chapter, the principals reported that for all five areas surveyed the domain that was most impacted was instruction. This was true for the overall responses, in a range from 75.0% to 82.9%, and for the target population in a range from 71.7% to 83.8%. Similarly, professional responsibilities were, without exception, the lowest impacted domain for the target population. For overall responses the range was from 47.9% to 53.5% and for the target population the range was from 37.2% to 45.7% that indicated that the four-tier teacher evaluation system impacted the various areas. The principals were permitted to select multiple categories which was why the total responses exceed 100%. Also important to note is that the number of responses

decreased from one question to the next as there was regression in participation as the study reached its latter questions.

Table 11

*Data Analysis Framework under Four-Tier Teacher Evaluation System for All Responses*

	Planning and Preparation	Classroom Environment	Instruction	Professional Responsibilities	None
Teacher Recognition (n=640)	58.0%	50.9%	76.9%	51.6%	10.8%
Teacher Effectiveness (n=620)	58.7%	53.5%	77.3%	47.9%	11.0%
Teacher Growth (n=609)	64.2%	55.3%	78.7%	52.2%	10.2%
Teacher Remediation (n=591)	67.0%	64.6%	82.9%	53.5%	10.3%
Teacher Dismissal (n=583)	57.3%	60.7%	75.0%	52.3%	18.4%

### Summary

Through the presentation of data displayed in this chapter the researcher has provided data which can be further analyzed and interpreted in the following chapter. The collection and presentation of data related to the survey administration and demographic information in concert with the principals' perceptions related to teacher recognition, effectiveness, growth, remediation and dismissal provided a detailed picture of the intended and unintended impact of the expanded teacher rating system in Florida and Massachusetts.

Table 12

*Data Analysis Framework under Four-Tier Teacher Evaluation System for Target Population*

	Planning and Preparation	Classroom Environment	Instruction	Professional Responsibilities	None
Teacher Recognition (n=186)	55.4%	48.9%	76.9%	45.7%	10.2%
Teacher Effectiveness (n=180)	53.9%	47.8%	78.9%	37.2%	10.0%
Teacher Growth (n=178)	64.0%	52.8%	77.5%	41.0%	8.4%
Teacher Remediation (n=173)	63.0%	57.2%	83.8%	39.9%	9.2%
Teacher Dismissal (n=173)	49.7%	53.8%	71.7%	43.4%	21.4%

In summary it was found that the majority of principals surveyed felt there was a need for multiple tiers to recognize standard attainment and standard deficiencies. By a slight margin the target population less preferred the four tiers in comparison to respondents overall. This margin was from 84.4% to 82.7% for attainment of standards and from 81.6% to 76.1% for deficiency standards. The intended impacts of the expanded four-tier system according to the principals included promoting teacher growth, promoting remediation, and supporting the dismissal of ineffective teachers. The unintended impacts of the expanded four-tier rating system included low morale, teacher stress, interfering with teacher growth, and making it harder to dismiss teachers. The principals in the target population also experienced less challenges to their assigned

rating in their prior system compared to their expanded rating system. The target population indicated that more teachers have been put under remediation in the expanded rating system and less teachers have been dismissed which may be a function of the amount of time evaluators have been able to use the new expanded performance rating system.

Placing these data in the conceptual framework of Charlotte Danielson's domains for effective teaching organized and displayed these data in a format that can be further interpreted and analyzed in the following chapter in order to answer the research questions that guide this study. It was found that instruction was the area most impacted by the expanded rating system for all areas studied (teacher recognition, effectiveness, growth, remediation, and dismissal). It was found that the least impacted area for all areas studied was professional responsibilities.

## CHAPTER V

### ANALYSIS OF DATA

#### **Introduction**

The analysis of data in this chapter is framed around conclusions based on the research questions that served as the guiding compass for this study. The data analysis of the research questions is followed by a discussion of the limitations of this study and the recommendations for further research based on this study. Lastly, a summary of the findings and the implications of these findings on the field of educational leadership is shared.

To review, the research questions of this study focused on the various perceived impacts, intended and unintended, created by the state mandated expansion to a four-tier teacher performance evaluation rating model. In order to capture these perceived impacts the following research questions were researched and answered:

From the perspectives of Massachusetts and Florida principals:

- 1) What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 2) What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?

- 3) What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 4) What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- 5) What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

## **Conclusions**

### **Research Question 1**

What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?

Principals in Florida and Massachusetts believed that the intended impact of expanding the teacher performance ratings in their respective states was primarily to promote teacher growth as was illustrated by the frequency of responses to the open-ended questions on this topic (see Table 13). In relation to teacher recognition the most frequent responses were to promote teacher growth and to recognize teacher excellence.

The same theme of promoting teacher growth was the most frequent response by principals in regards to teacher effectiveness and, naturally, teacher growth as well. A similar theme emerged in regards to the intended impact of the expanded rating system on teacher remediation; this theme was promoting teacher growth and remediation. Lastly, the most frequent response for the intended impact of the expanded performance rating system on teacher dismissal was to support the dismissal of ineffective teachers.

Table 13

*Summary of the Intended Impacts of the Four-Tier Rating System*

<b>Theme</b>	<b>Most Frequent Response(s)</b>	<b>Percentage</b>
Teacher Recognition (n=154)	Promote Teacher Growth Recognize Excellence	32.5% 26.0%
Teacher Effectiveness (n=119)	Promote Teacher Growth	37.0%
Teacher Growth (n=118)	Promote Teacher Growth	47.5%
Teacher Remediation (n=104)	Promote Teacher Growth/ Remediation	42.9%
Teacher Dismissal (n=98)	Supports Dismissal of Ineffective	44.9%

The concept that teacher evaluation should promote teacher growth as a primary function is a widely held belief (Danielson, 2013; Fullan, 2010; Israel & Kersten, 2007; Marzano et al., 2011; Sergiovanni, 2007; Stronge et al., 2006; Zepeda, 2007). As Stronge, Gareis, and Little (2006) stated, the “primary goal of a teacher evaluation system should be to encourage continuous growth and improvement at an individualized level by collecting and analyzing pertinent data and utilizing those data as a foundation for meaningful feedback” (p. 28). Massachusetts and Florida made this promotion of growth connection explicit with their recent teacher evaluation legislation (Florida Department of

Education, 2013; U.S. Department of Education, 2012) which explains why the preponderance of principals reported the promotion of teacher growth was the intended impact of their respective expanded performance rating systems.

### **Research Question 2**

What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?

Principals in Florida and Massachusetts believed that there were various unintended impacts in regards to expanding the teacher performance ratings in their respective states. The lower frequency level of responses for each question, in comparison to the previous research question focused on intended impacts, may indicate that it was clear what the state intended to do but the actual outcomes were different from what the state intended and were multiple in nature. For example, the principals in the study indicated that the *intended* impact of the expanded performance rating system was to promote teacher growth but in actuality the *unintended* result was low morale and interference with teacher growth.

The themes of low morale and interfering with growth were two recurring themes in regards to the unintended impacts of the expanded performance rating system (see Table 14). Low morale surfaced as a frequent response in relation to the unintended impacts of teacher recognition, teacher effectiveness, and teacher growth. The theme of the interference with growth as an unintended impact appeared in regards to teacher recognition, teacher effectiveness, teacher growth, and teacher remediation. Teacher stress was the most frequent response in relation to teacher effectiveness.

Table 14

*Summary of the Unintended Impacts of the Four-Tier Rating System*

<b>Theme</b>	<b>Most Frequent</b>	<b>Percentage</b>
Teacher Recognition (n=154)	Low Morale Interferes with Growth	16.9% 13.0%
Teacher Effectiveness (n=121)	Teacher Stress Low Morale Unfair Interferes with Growth	11.6% 8.3% 8.3% 8.3%
Teacher Growth (n=107)	Interferes with Growth Promotes Teacher Growth Low Morale	15.9% 9.3% 8.4%
Teacher Remediation (n=101)	Achieved Intended Impact Interferes with Growth Teacher Attrition	15.8% 9.9% 8.9%
Teacher Dismissal (n=91)	No Unintended Impact Harder to Dismiss Teachers Union	20.9% 12.1% 11.0%

The responses for teacher remediation and teacher dismissal indicated that there was alignment between the intended and unintended impacts. The most frequent response under the areas of teacher remediation was that the expanded performance rating system achieved its intended impact and the most frequent response for teacher dismissal was that there were no unintended impacts.

The themes of low morale and teacher stress as unintended impacts should not come as a total surprise with a new evaluation system and its accompanying expanded performance ratings. Danielson reported in 2007 that “An environment of high-stakes accountability only exacerbates teachers’ levels of stress... Teachers are under enormous external pressure, as never before, to prepare their students for productive lives in the knowledge economy and success in externally mandated assessments” (p. 5).

The performance evaluation rating system could even prove to be stressful and unproductive for those that are achieving at the highest levels due to the emphasis placed on areas such as excellent ratings and/or merit pay. Turning to Daniel Pink's *Drive* (2009), he cites that extrinsic rewards, "in particular, contingent, expected, 'if-then' rewards" discouraged drive and motivation (p. 37). Instead, Pink argues that people "*want* to be accountable – and that making sure they have control over their task, their time, their technique, and their team is the most effective pathway to that destination" (p. 105).

Curiously, the interference with growth and the promotion of growth are the top two themes that emerged in regards to the unintended impacts on teacher growth. As was cited by one principal respondent in regards to interference with growth, "Some teachers just aren't ready for change and so they get caught up on the rating instead of how they can grow as a professional and educator." In contrast, another principal respondent believed an unintended consequence was the further promotion of growth beyond what was intended by the state. This respondent shared in the survey that, "Teachers can't hide behind something that they did years ago – they need to be constantly making improvements. This is a good thing." Either way, growth is valued in these comments and the role of the principal as the instructional leader may be the link between these varying views and this concept will be further explored later in this chapter.

Another unintended impact that was cited under the area of teacher remediation was the theme of teacher attrition. Although this may at first be perceived as a negative, after taking a closer look at the responses it was actually viewed as both a positive and negative unintended outcome. Some of the positive outcomes cited included:

- “Teachers who are ‘burned out’ have left the profession.”
- “The teachers that fall into the unsatisfactory, or close to that range, shut down. Some quit teaching.”
- Some teachers may never change. They retire early.”

Conversely, some principals cited the attrition of teachers as a negative:

- “Some good even great teachers will walk.”
- “Qualified potential candidates choose a different profession.”

Although the new evaluation system may be getting the poor teachers out through attrition it may also be discouraging effective educators from remaining in the field.

Teacher attrition is already a challenge to the educational profession to the detriment of continuous improvement efforts nationwide. It has been found that attrition rates of nearly 50% exist for teachers in the first five years of their profession (Metlife, 2009). A study conducted by the Institute of Education Sciences found that teachers who left their teaching position cited that the procedures for performance evaluation were two times better in their new professional position (28.9%) compared to their prior position in teaching (14.6%) (Keicher, 2010).

In relation to teacher attrition the question remains whether the ineffective teachers are leaving or the effective teachers are leaving. If it is indeed the ineffective teachers who are leaving this could be a positive. If these unsatisfactory teachers are leaving on their own due to the rigors and clear communication provided by the evaluation process this is saving administrators’ time and getting poor teachers away from students. On the other hand, if the evaluation process is discouraging the best

teachers from remaining in the profession and in turn leaving students in the classroom with less successful teachers then the educational system has a problem on their hands.

Also curious was the unintended theme of the four-tier rating system making it harder to dismiss teachers, which conflicts with the intent of both Massachusetts and Florida (Massachusetts Department of Elementary and Secondary Education, 2011; National Council on Teacher Quality, 2011). Although 44.9% of respondents cited that an intended consequence was to support the dismissal of teachers there were 12.1% of principals that felt an unintended consequence was that it became harder to dismiss teachers. Survey comments from principals that captured the theme of making it harder to dismiss teachers included:

- “Because their final evaluation is a combination of the instructional framework observation and the student performance (VAM) which is muddy, it isn't as effective. Only time will tell.”
- “In some cases it becomes more complicated since teachers are still new to the system.”
- “We are currently struggling to provide organized district support to all the teachers who “need improvement,” so we have limited to unsatisfactory or teachers who are multiple years “N I”. We have NOT been provided adequate budget to support this system!”

The theme of time to implement and understand the model emerges here and will be further built upon later in this chapter.

The concept of union pushback and involvement is another significant finding within the unintended consequences of the teacher dismissal process under the four-tier

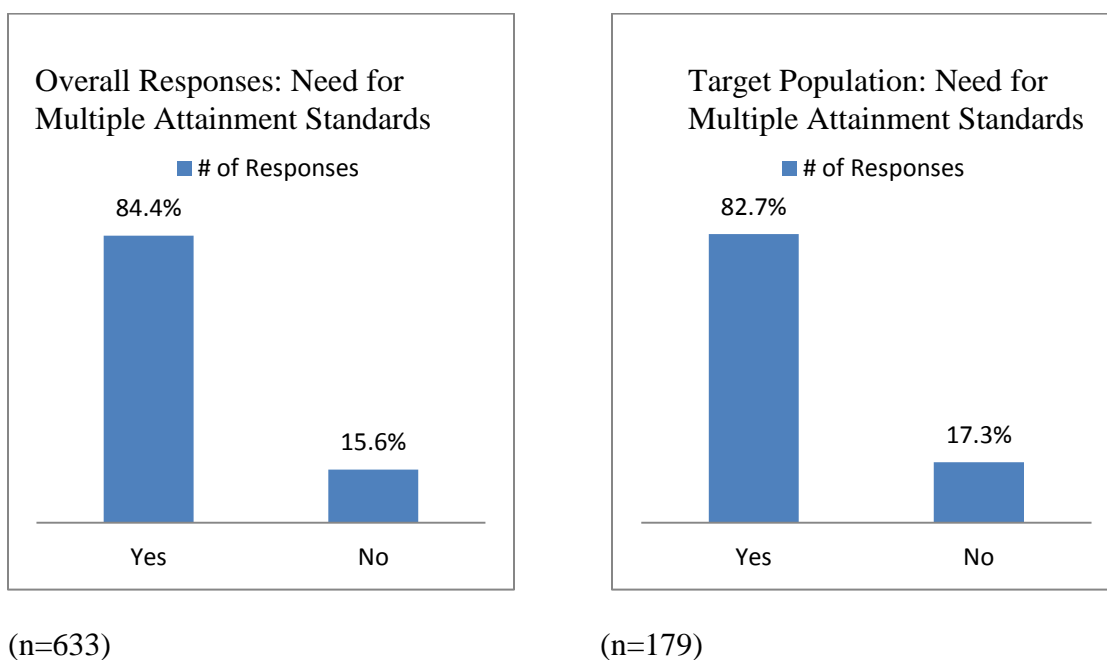
rating system. This confirms what Kersten and Israel (2005) warned about which was that “principals perceive unions as not trusting the more complex, subjective teacher evaluation methods that are currently considered best practice” (p. 62).

### **Research Question 3**

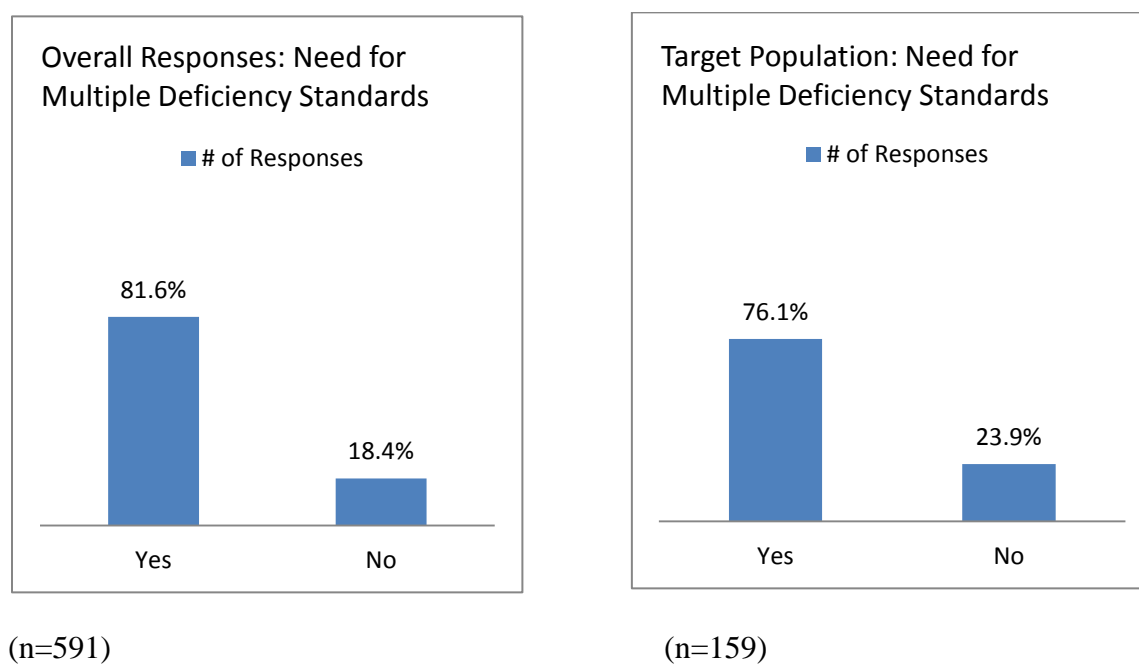
What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., “Excellent,” “Proficient”) or standard deficiency (i.e., “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?

As presented in Chapter IV, the overall responses and the target population responses both showed the vast majority of principals believed there was a need for multiple standards to recognize teacher standard attainment (see Figures 42 and 43). The most frequent response to explain why the multiple tiers were necessary was to delineate performance. This was the most frequent answer for recognizing standard attainment (30.0%) and for recognizing standard deficiency (42.6%).

A similar trend was found in regards to the need for a performance rating system with more than one tier to identify standard deficiency (see Figures 42 and 43). To further reinforce the preference of the principals for multiple ratings it was found that 86.3% of the overall responses (n=666) and 80.3% of the target population (n=187) preferred four or more teacher performance rating categories. This confirms the argument presented in *Teacher Evaluation 2.0* (2010) which posited that a four or five tier rating scale “is large enough to give teachers a clear picture of their current performance, but small enough to allow for clear, consistent distinctions between each level” (NCTQ, p. 7).



*Figure 42. Necessity for a System with More Than One Tier to Recognize Teacher Standard Attainment*



*Figure 43. Necessity for a System with More Than One Tier to Identify Standard Deficiency*

As was acknowledged in Chapter III, the researcher may have had a personal bias in regards to different teacher evaluation rating systems. The researcher has personally evaluated teachers in a binary system, a three tiered system, and a four tiered system of performance evaluation ratings. The researcher did not expect such a conclusive preference for multiple tiers to identify standard attainment and standard deficiencies based on his own biases. The quantitative nature of the study, along with a running critical reflection kept in the researcher's journal, minimized this bias and allowed this critical conclusion to emerge.

#### **Research Question 4**

What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e., "Excellent," "Proficient") or standard deficiency (i.e., "Needs Improvement," "Unsatisfactory") on teacher recognition, effectiveness, growth, remediation, and dismissal?

The data did not reveal any significant unintended impacts from the multiple teacher performance ratings for standard attainment or standard deficiency. In hindsight this research question could have been asked more directly and this point will be raised later in this chapter under recommendations for further research.

The one response that came up with some frequency which may support the existence of unintended impacts was that the multiple standards were irrelevant. In regards to standard attainment this answer was indicated in 5.8% of the target population responses. In regards to standard deficiency this answer was indicated in 12.0% of the target population responses.

Principal comments related to the theme of the multiple standards being irrelevant included the following:

- “Professional growth would occur within a more supportive and collaborative climate if teachers were not vying for the highest rating within a tiered system.”
- “System is really meant to identify poor teachers not proficiency.”
- “We are complicating what should be a fairly simple system. If all of the angst does not result in improved student performance, then why is it necessary?”
- “A system with on tier can identify deficiencies. Simply – the teacher is either meeting the standard or not meeting the standard. It is black and white.”
- “Evaluation system could work if used with fidelity, but it is not.”

The absence of any significant evidence pointing to the unintended impacts of multiple teacher performance ratings for standard attainment and deficiency may also logically indicate that the multiple tiers met their intended impact. This would be another explanation for the limited amount of data compiled on this specific research question.

### **Research Question 5**

What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

As the State of Illinois travels down their uncharted road of expanding to a four-tier teacher performance rating evaluation model they can learn much from Florida and Massachusetts who recently trail blazed the same road themselves. Instruction was conclusively the most impacted domain while the domain of professional responsibilities

was clearly the least impacted domain. The study data also revealed a need for more time for principals; both within the academic year to meet the significant teacher evaluation demands and across academic years to more effectively implement the model. More teachers were on remediation while fewer teachers were being dismissed. Principals reported there were more challenges to the assigned teacher evaluation rating given but these principals still asserted the performance rating given did promote growth.

### **Instruction as the Most Impacted Domain**

There is much that Illinois educational leaders can learn from the experiences of Florida and Massachusetts in regards to their recently expanded performance rating system. One of the most pronounced conclusions was that instruction was the domain most impacted for each of the areas surveyed (teacher recognition, effectiveness, growth, remediation, and dismissal). This conclusion served as the roadmap for the study findings as it was true for the overall responses, in a range from 75.0% to 82.9%, and for the target population in a range from 71.7% to 83.8%. In Chapter IV, the data analysis for this study's conceptual framework of the Danielson's domains was juxtaposed with the elements of teacher recognition, effectiveness, growth, remediation, and dismissal. This comparison also follows on the next page with instruction bolded for emphasis as the most frequent response (see Tables 15 and 16). The principals' perceptions in the study consistently indicated across all questions that instruction superseded planning and preparation, classroom environment, and professional responsibilities. This finding communicated that instruction was the most impacted domain in the expanded performance rating system. Given the importance of instruction, as reported by the principals surveyed, the concept of the principal as the instructional leader and as the

lever for improving student achievement becomes an important one. This will be explored further later in this chapter under significance for educational leadership.

Table 15

*Data Analysis Framework under Four-Tier Teacher Evaluation System for All Responses*

	Planning and Preparation	Classroom Environment	<b>Instruction</b>	<i>Professional Responsibilities</i>	None
Teacher Recognition (n=640)	58.0%	50.9%	<b>76.9%</b>	51.6%	10.8%
Teacher Effectiveness (n=620)	58.7%	53.5%	<b>77.3%</b>	47.9%	11.0%
Teacher Growth (n=609)	64.2%	55.3%	<b>78.7%</b>	52.2%	10.2%
Teacher Remediation (n=591)	67.0%	64.6%	<b>82.9%</b>	53.5%	10.3%
Teacher Dismissal (n=583)	57.3%	60.7%	<b>75.0%</b>	52.3%	18.4%

The significant impact of the expanded four-tier rating system on instruction reinforces the position of Charlotte Danielson (2007) whose framework for teaching served as the conceptual framework for this study. She stated that her domain three of instruction was “the heart of the framework for teaching; it describes, after all, the critical interactive work that teachers undertake when they bring complex content to life for their students” (p. 77). Furthermore, Danielson ascertained that the component of engaging students was the heart of instruction as “all the other aspects of the framework serve the purpose of engagement, because it is engagement that ensures learning” (p. 77). This also reinforces research that indicates that effective teachers and effective instruction are

the most important elements to promoting student achievement (Marzano et al., 2011; Schmoker, 2011; Toch & Rothman, 2008).

Table 16

*Data Analysis Framework under Four-Tier Teacher Evaluation System for Target Population*

	Planning and Preparation	Classroom Environment	<b>Instruction</b>	<i>Professional Responsibilities</i>	None
Teacher Recognition (n=186)	55.4%	48.9%	<b>76.9%</b>	<i>45.7%</i>	10.2%
Teacher Effectiveness (n=180)	53.9%	47.8%	<b>78.9%</b>	<i>37.2%</i>	10.0%
Teacher Growth (n=178)	64.0%	52.8%	<b>77.5%</b>	<i>41.0%</i>	8.4%
Teacher Remediation (n=173)	63.0%	57.2%	<b>83.8%</b>	<i>39.9%</i>	9.2%
Teacher Dismissal (n=173)	49.7%	53.8%	<b>71.7%</b>	<i>43.4%</i>	21.4%

**Professional Responsibilities as Least Impacted Domain**

Another conclusive finding was that the domain of professional responsibilities was the least impacted domain in the expanded teacher performance rating system. This was true for all of the areas in the target population and for all but one area in the overall responses (teacher recognition). This finding regarding professional responsibilities is italicized for emphasis in all the areas it received the lowest number of responses (see Tables 15 and 16 on the previous pages).

The relatively low impact that professional responsibilities had with the expanded performance ratings likely comes as little surprise to Danielson (2007) who commented on the nature of professionalism by stating that “Teaching has been treated – and, to some degree, has treated itself – as a job, with almost an assembly-line mentality, in which teachers follow a ‘script’ that has been designed by someone else, presumably more expert” (p. 18). The underscoring of professional responsibilities flies in the face of literature that emphasizes the critical role that professional responsibilities, and in turn teacher growth and development, play in an effective teacher evaluation system (Israel & Kersten, 2007; Sergiovanni, 2007; Stronge et al., 2006; Zepeda, 2007).

There are efforts around the nation to address the lack of emphasis on the importance of professional responsibilities, growth, and development for both teachers and leaders. One effort that is in progress to address the relative deficiency around professional responsibilities is the “Principal Pipeline” spearheaded by The Wallace Foundation. This foundation is embarking on a \$75 million initiative in six urban districts and will focus on the key elements of the principal pipeline which their research has identified as rigorous job requirements, high-quality training, selective hiring, and on-the-job evaluation and support (The Wallace Foundation, 2011). The KIPP Foundation is another example of emphasizing the professional growth of principals to promote the effectiveness of teachers and in turn the success of students. This foundation has established school leadership programs which recruit and train leaders to open and operate their schools (The KIPP Foundation, 2013). These foundations are two examples which offer insight regarding what needs to change in order to further emphasize the importance of professional responsibilities in the educational field.

### **Need for More Time**

Another theme that emerged from the study was that the states of Florida and Massachusetts need more time. The need for more time applied both to evaluating teachers within each academic year and also from one year to the next to further implement the new model, before the effectiveness and impact of the expanded performance rating system can be fully measured.

The teacher evaluators that were part of the target population indicated they had a significant evaluation load; 92.0% of principals indicated they observed sixteen or more teachers in a year. Combining these significant evaluation loads with the amount of time it takes to effectively evaluate teachers leads to problematic results, as some responding principals indicated:

- “It is not doable for administrators. The system (paperwork) is so cumbersome for administrators that it will end up being as ineffective as the old system; whereas only use to target underperforming teachers and mediocre teachers will continue to get by.”
- “It is difficult to keep up with the volume and length of evaluations so some evaluators tend not to give needs improvement.”
- “Principals spend the majority of their day doing some aspect of teacher evaluation.”
- “Not enough time to teach the teachers all the information about indicators, etc. that they need to know to understand how they are being evaluated.”

One important question principals need to ask themselves is whether the problem is the amount of time the evaluation process takes or if the problem is that principals

typically spend their time in other areas and need to make a shift in their priorities and present operations. Recent studies have reported that the average principal dedicates around 18% of their time to the area of instruction and curriculum and approximately 3% on teacher evaluation (May & Supovitz, 2011). Murphy, Hallinger, and Heck (2013) assert that even if principals were willing to dedicate more time to the evaluation process those efforts “overlook the existing evidence on the willingness of the public to increase administrative costs in education” (p. 351). Instead, Murphy et al. argue that school administrators are better off going in a different direction as “Studies also tell us that school administrators will be more likely to positively impact instructional quality if they allocate their direct efforts with teachers into facilitative channels” (p. 352).

This conflict between having enough time and deciding where to spend time is an important one. In reality this researcher would argue from his own life experiences that much of the emphasis should be placed on where we choose to spend our time as educational leaders. The researcher chose to make teacher evaluation and instruction a priority over the past three years since coming to the Illinois education system. Over these past three years the researcher has started a new administrative job in a new state, had two new children, bought a new house, started and completed the coursework and dissertation for a doctoral program, and stepped into his first principal job a year ago. Despite these demands on time the researcher would argue his dedication to evaluation and instruction exceeded the averages found in the previous paragraph because that is where time was intentionally allocated and prioritized.

An argument could be made that the existence of more time, in regards to years of experience in the new model, might provide principals the opportunity to learn how to

put more teachers on remediation. In analyzing the overall responses to the survey, which included those with less than two years of experience in the expanded rating system, 32.1% of principals had put two or more teachers per year on remediation with the new system in comparison to the target population which indicated that 43.4% of principals had put two or more teachers on remediation per year. This indicated that the expanded teacher performance rating system has led to more teachers being placed on remediation as will be discussed in the next section.

### **More Teachers on Remediation**

The principals in the target population also indicated that they now put more teachers on remediation in their expanded four-tier rating system in comparison to their prior evaluation system. In the current four-tier evaluation system 43.4% of principals in the target population indicated they put two or more teachers on remediation per year compared to 30.8% of principals in the target population that said they put two or more teachers on remediation per year in their prior evaluation system. This may indicate or suggest that some of these teachers presently on remediation may soon move towards dismissal as time takes its course with the new model. However, it is possible that remediation provides a road for improvement which would lead to less teachers being dismissed; this concept will be further analyzed in the following section.

To further clarify, once a teacher enters into remediation they have only two potential outcomes, remediation and growth which will place them back in a satisfactory standing or a failure to remediate which would result in dismissal. This could also explain why the principals in the study indicated a drop in dismissals with the recently expanded four tier performance rating system. Given the increased number of teachers

being placed on remediation in the expanded performance rating system it is logical to believe that more of these teachers will eventually be dismissed at the end of their remediation cycle in comparison to the prior evaluation system which reported less teachers having been on remediation.

### **Less Teachers Being Dismissed**

The target population that has lived the extended four-tier model for two or more years interestingly had more difficulty terminating ineffective teachers. As shared in Chapter IV (see Figure 13), when asked to compare to their prior system the target population indicated that they were less effectively able to terminate ineffective teachers (46%) in comparison to the overall principal responses for the same question (55%).

Although the principals in the target population did indicate they are now putting more teachers on remediation in comparison to their prior system they did not indicate that they are terminating more teachers and this may also be a component of time. In their prior evaluation system 45.3% of principals indicated they dismissed one or more teachers per year compared to 26.7% of principals using the expanded four-tier system. One could argue that teachers are getting better and this expanded performance rating system is effectively promoting growth. Given the teacher dismissal information presented one could also speculate that more teachers will be terminated as the expanded four-tier system has had more time to be implemented; as was just presented in relation to teacher remediation which has shown that more teachers are under remediation in the newly expanded performance rating model.

The theme of time does not only relate to the principals' ability to put teachers on remediation and/or terminate them but it also surfaced in relation to teachers growing in

order to avoid termination. As one principal related in regards to promoting growth of a teacher on remediation, we need “to enable new teachers time to grow without penalty, veteran personnel to adapt to changes, and ultimately everyone focus on continuous improvement.” As Nolan and Hoover (2005) argue, “Remediation as a goal makes sense from a variety of viewpoints – ethical, organizational, legal, and economic...the primary reason for investing time and effort to remediate the performance of marginal teachers is thus moral and ethical. It is simply the right thing to do” (p. 300). Jackson (2008) supports this same argument by stating that “Yes, it is important to eliminate mediocre or poor teaching, but the best way to get rid of mediocre or poor teaching is to help those teachers improve” and this is done by focusing on a culture of growth and improvement with appropriate support to improve teacher practice (p. 9).

### **Analysis of Teacher Rating Protests and Rating Interference with Growth**

When principals in the target population were asked how many challenges or protests they encountered by teachers under the four-tier system and under their prior system the resulting data revealed a significant difference. In the expanded four-tiered system, 69.8% of principals experienced one or more challenges or protests per year to the performance ratings they assigned teachers. This compares to 38.3% of principals in the target population that had one or more challenge or protest per year in their prior system. This could be a function of time as well and the number of challenges may decrease as everyone becomes familiar with the new system. As the principals are better able to use the new expanded performance ratings and their accompanying structures there will likely be more consistency in its effective implementation and usage which may in turn reduce the number of protests. However, another perspective is that the

expanded performance ratings instead could have established more opportunities for differing viewpoints and delineations of teachers' performance resulting in push back in comparison to the prior model with less performance ratings.

Although principals reported that they are encountering a greater number of challenges in the expanded four-tier rating system few principals assert that the impact of the rating serves to distract from teacher growth. In fact, only 5.6% of the target population reported that the performance rating the teacher received distracted from growth while an overwhelming 77.0% of principals reported that the rating promoted growth. This also contradicted a personal bias of the researcher who believed that the expanded ratings would indeed further distract from the growth of the teacher which should be the ultimate goal of the evaluation process. This was captured in the researcher's dissertation journal which documented the feeling of surprise by the researcher at the low number reporting that the expanded rating interfered with teacher growth. This low number reporting interference of growth contrasted with the personal bias and prediction of the researcher.

The high number of principals reporting the promotion of growth through the ratings given may be due to the clear and common language built around the evaluation by using frameworks such as those developed by Marzano, Danielson, and others. As Danielson (2007) stated, "During conversations about practice, particularly when such conversations are organized around a common framework, teachers are able to learn from one another and to thereby enrich their own teaching" (p. 6).

In further focusing on growth, 82.6% of principals in the target population acknowledged that the performance rating had "some" or a "significant" impact on the

promotion of the teacher growth areas they identified in the evaluation. Breaking this number down further it was found that 71.3% of the 82.6% of principals believed there was only some impact. This result is not necessarily a bad thing but given the amount of time dedicated to the evaluation process an evaluator would desire a greater outcome in regards to the promotion of growth. This idea of dedicating the appropriate time to professional growth in order to attain desirable outcomes was reflected in the work of Israel and Kersten (2007) who found “More and more, educators are recognizing that the staff development plans must be multi-dimensional, sustained over time, and include opportunities for faculty members to learn, discuss, experiment, and apply new knowledge within their individual classroom with adequate support” (p. 55).

### **Teacher Morale and Stress are Recurring Unintended Themes**

Two of the many significant unintended impacts of the expanded performance rating system were low teacher morale and increased teacher stress. Low morale surfaced as a frequent response in relation to the unintended impacts of teacher recognition, teacher effectiveness, and teacher growth. Teacher stress surfaced in relation to teacher recognition, teacher effectiveness, teacher growth, and teacher dismissal.

Revisiting Chapter IV, there were a number of principal comments directly related to this topic:

- “Teacher morale has been very low with the new four tier rating system as the perception is that the new system is an effort to rate more teachers as effective or needs improvement”
- “The state has intentionally made it difficult for teachers to receive ratings of excellent. The belief is that proficient is what people should aspire to and that

excellent should be rare. It creates a demoralizing dynamic for excellent teachers.”

- “The unintended impact is an atmosphere of mistrust and negative feelings on the part of teachers who have always received the highest level rating under the three tier system and now are not.”
- “Teachers have become quite sensitive to the rating system and its impact on their careers and finances. Teachers who view themselves as highly effective, but are not evaluated at that level become frustrated and demoralized.”

Given the wide reaching impact that morale and stress are having on the themes studied, from teacher recognition to teacher dismissal, its importance for principals to understand and address is critical. Sergiovanni (2007) offers some advice on how to address this concern by promoting an effective climate and culture around teacher evaluation by stating, “Much of the discomfort concerning evaluation can be eliminated, however, if it is treated as a community exercise in self-governance, as a way for the school community to maintain and strengthen its commitment to learning” (p. 297). This complements another perspective by Toch and Rothman (2008) referenced in chapter two of this study which state that components such as scoring rubrics, multiple classroom visits by multiple evaluators, student work, and teacher reflections “contribute much more to the improvement of teaching than today’s drive-by evaluations or test scores alone. And they contribute to a much more professional atmosphere in schools” (p. 13).

### **Limitations of the Study**

In analyzing the responses it was not found that a response bias of any kind existed in the research data. The respondents represented a variety of demographics that

one would likely find in a survey of principals. The researcher does not believe that non-respondents could have substantially impacted the overall results in any way given the demographic distribution that was observed in the study.

There were other limitations that could have impacted the study in various ways. The researcher did not actually talk to any of the respondents as the survey was conducted electronically. There was no teacher voice captured in the research and this demographic certainly possesses a wealth of information that could further inform the research. There was also no data collected to analyze if there was a relationship between the expanded teacher evaluation ratings and its impact on student achievement. The logical correlation and conclusion would be that if teachers are growing then student achievement should be improving.

The demographic data did uncover a fact that was not discovered during the researcher's literature review. It was publicly communicated that both Massachusetts and Florida were establishing rating systems with four tiers beginning with the 2011-12 academic year. However, in this study a relatively small number of Massachusetts principals met the criteria necessary to be identified as part of the target population. Instead, Massachusetts saw a big jump in the number of principals that were using an expanded four-tier rating system for two or more years in the 2012-13 year. What was not discovered by the researcher prior to the survey administration was that although Massachusetts did adopt a four-tier system in 2011-12 it was only implemented in thirty-four schools along with fifteen early adopter districts. In 2012-13 all Massachusetts Race to the Top districts would implement the four-tier performance rating system (MDESE, 2011). Given the underrepresentation of the Massachusetts principals in the target

population it is possible that Massachusetts principals could have influenced the data differently had it been better represented.

### **Recommendations for Future Research**

This study could open the door for additional research to be done in a number of areas. One area would be an expanded version of the study which would analyze other states with similar models to Illinois. This would help to affirm and bring further credibility to the findings of this research or perhaps refute the results of this research should different findings emerge. Another area would be to study states that employ a completely different model to that of the states studied in this dissertation. By using the same research questions to study states with different models from Illinois the results could potentially find more effective systems that should be considered for implementation.

A significant future study could be to conduct an in-depth analysis of the Illinois four tier model after it has had sufficient time to be implemented. By doing so, principals in Illinois would have the same opportunities as the Florida and Massachusetts principals in this study to provide their insights regarding the effectiveness of the expanded teacher performance rating system. It may also be worthwhile to engage in a teacher perception study to juxtapose their views with those of the principals surveyed in this study. Teachers did not have a voice in this research and their perceptions regarding the research questions could provide significant insight which could be compared to the principals' perceptions in measuring the effectiveness of the expanded model.

All of these recommendations could potentially incorporate a case study component that directly talks to the principals, or teachers, to provide a more in-depth

analysis of the research questions with a greater ability to ask follow-up questions. All of these recommended areas would provide additional perspectives and data that could either confirm or refute the conclusions of this research.

Further analysis could be considered regarding the unintended impact of the four-tier rating system in relation to teacher growth where the most frequent responses conflicted with one another; the interference with teacher growth and the promotion of teacher growth. Given the polar opposites of these two responses further analysis may provide additional insight regarding the impact of the expanded performance ratings.

The concept of merit pay was not directly questioned in this study, yet it was an important component that was cited in the open responses by the principals. Further research could be conducted on the role of merit pay in promoting teacher effectiveness in order to further inform its impact on the research questions. For example, the role merit pay played in the need for multiple tiers to identify standard attainment could be asked more directly to tease out the role it played in the responses.

Additional areas of focus for future study could also include an analysis of whether or not more teachers will be dismissed per year as the expanded four-tier system is implemented for additional years. This would further answer the questions and discussion related to the need for more time to determine the impact of the expanded teacher performance rating system that was posited in this chapter. In addition, further study could be conducted regarding the increased number of challenges and protests to the performance rating in the expanded model and whether this was a function of a new system or a sustained change in the teacher/evaluator dynamic. Another important future recommendation would be to analyze data related to the correlation between the

expanded teacher evaluation ratings and their impact on student achievement. This specific question is probably the most important of all of the potential future research questions given that student achievement lies at the heart of all of the work that educators engage in every day.

Lastly, the third and fourth research question in this study could have been asked more directly in order to extract more specific data on their respective topics. These two questions were related to the intended and unintended impacts of providing multiple teacher performance ratings to recognize standard attainment and standard deficiency. Quantitative data was collected around these two areas but little was collected specifically in relation to qualitative and anecdotal data. The principals in this study were asked to explain their answers regarding the need for multiple standards to demonstrate standard attainment or deficiency but explicit follow-up questions related to the intended and unintended impacts of these multiple tiers could add additional value to these research questions.

### **Significance to Educational Leadership Practice and Preparation**

This study has a number of implications for educational leadership practice and preparation. Two major takeaways that were found in the research were the element of time and the critical importance of the principal as the instructional lever for creating improved student outcomes.

### **Time Element**

Time is an integral component in regards to learning more about the potential that the expanded performance rating systems and the effectiveness of the rating system to impact teachers. The lesson that Illinois can learn from this study is that more time is

needed to learn the new system and allow principals and teachers to adapt to the system. Illinois would be wise to continue to learn not only from their own experiences in their expanded four-tier model but also to continue to learn from those states that are out ahead of them trailblazing the way. Furthermore, Illinois should sufficiently measure outcomes before changing their course in any way.

System reform such as what was studied in Florida and Massachusetts will face adversity and pushback as change often does. If time is what is needed in order to prove effectiveness then sound leadership is the key to making a commitment to that time. As Michael Fullan (2010) states, “Resolute leadership is critical near the beginning when new ideas encounter serious difficulty, but it is also required to sustain and build on success” (p. 5). Fullan’s work with whole system reform asserts that one of the primary components to effective reform is unyielding leadership that stays on message and “stays with the focus, especially during rough periods, and these leaders cause others around them to be resolute” (p. 5).

Jim Collins’ (2005), “Turning the Flywheel” analogy offers further leadership and systems research evidence that staying the course and committing to the model will produce positive outcomes. Collins argues that institutions should not look for an instantaneous, snap of the fingers, miracle fix, but instead:

Our research showed that it feels like turning a giant, heavy flywheel. Pushing with great effort – days, weeks and months of work, with almost imperceptible progress – you finally get the flywheel to inch forward. But you don’t stop. You keep pushing, and with persistent effort, you eventually get the flywheel to complete one entire turn. You don’t stop. You keep pushing, in an intelligent and consistent direction, and the flywheel moves a bit faster...Then, at some point – breakthrough! ...This is how you build greatness. (p. 23)

Given the increased number of teachers on remediation and the decreased number of teachers being dismissed under the expanded performance rating system this researcher would predict one of two things are occurring. The first possibility is that as the evaluators are provided more time to operate under this expanded performance rating system they will move more of these teachers from remediation to dismissal. This drive to ‘dismiss ineffective teachers’ would align with what was reported by the principals in this study as a significant intended impact of the new evaluation system by their respective states. The second possibility is a far more optimistic one which is that the teachers being identified for remediation are exhibiting growth and moving out of this category which in turn is decreasing the amount of teachers being terminated. This ‘promotion of growth’ would also align with the states’ intended impacts as reported by the principals in the study.

### **Principal as the Instructional Lever**

Given the consistent and clear emphasis on instruction that was found in this study the focus of the principal should then turn to their role as the instructional lever for continuous student improvement with management playing a secondary role. As the data from Massachusetts and Florida revealed, the principals’ perceptions consistently indicated across all questions that instruction superseded planning and preparation, classroom environment, and professional responsibilities which communicated that instruction was the most impacted domain in the expanded performance rating system.

Marzano et al. (2005) cite three specific instructionally related behaviors and characteristics that their meta-analysis found as important responsibilities of principals as instructional leaders:

- Being directly involved in helping teachers design curricular activities;
- Being directly involved in helping teachers address assessment issues;
- Being directly involved in helping teachers address instructional issues (p. 54).

Reeves (2004) also emphasizes the importance of the principal as an instructional leader by explaining that in an effective school “the principal personally evaluates student work and participates in collaborative scoring sessions...The principal personally reviews faculty-created assessments as part of each teacher evaluation and coaching meeting” (p. 50).

The principal as the instructional lever includes his or her ability to provide focused feedback to teachers. Ericsson, Krampe, and Tesch-Romer (1993) cited that, “In the absence of feedback, efficient learning is impossible and improvement only minimal even for highly motivated subjects” (p. 367). The concept of evaluator feedback was reinforced by Hattie and Timperley (2007) who found in their analysis of 12 meta-analyses of 196 studies that the average effect size for providing feedback was 0.79 which is approximately twice the effect size of most educational initiatives. Marzano et al. (2011) reinforced the importance of feedback and asserted that “For feedback to be instrumental in developing teacher expertise, it must focus on specific classroom strategies and behaviors” (p. 6).

Leithwood, Seashore Louis, Anderson, and Wahstrom (2004) found in their research that “Of all the factors that contribute to what students learn at school, present evidence led us to the conclusion that leadership is second in strength only to classroom instruction. Furthermore, effective leadership has the greatest impact in those

circumstances (e.g., schools “in trouble”) in which it is most needed” (p. 70). When Zepeda (2007) discussed the importance of the principal as the provider of cohesion between instructional programming and the vision and mission of the school the first thing she cited was the supervision of instruction (p. 6). She goes on to state that “Instructional leadership is an elusive concept; however, effective principals engage in work that supports teachers in improving their instructional practices, and this type of support occurs in the classrooms, not the principal’s office” (p. 10).

### **Summary of Findings**

In summarizing the findings (see Figure 44) the research questions that drove this study found that expanding the performance ratings had the intended impacts in Florida and Massachusetts of promoting teacher growth, recognizing teacher excellence, promoting remediation, and supporting the dismissal of ineffective teachers. This expansion of the performance ratings was also found to have unintended impacts such as creating low teacher morale, interfering with teacher growth, producing teacher stress, making it harder to dismiss teachers, increasing teacher attrition, and creating union pushback. In addition, the intended impacts of having multiple tiers to identify standard attainment and standard deficiency (see summary of research question #3) served to delineate performance between different levels of teacher performance. However, no unintended impact (see summary of research question #4) was found in the data which is in part due to the fact that this question should have been asked more directly which was addressed in recommendations for future research.

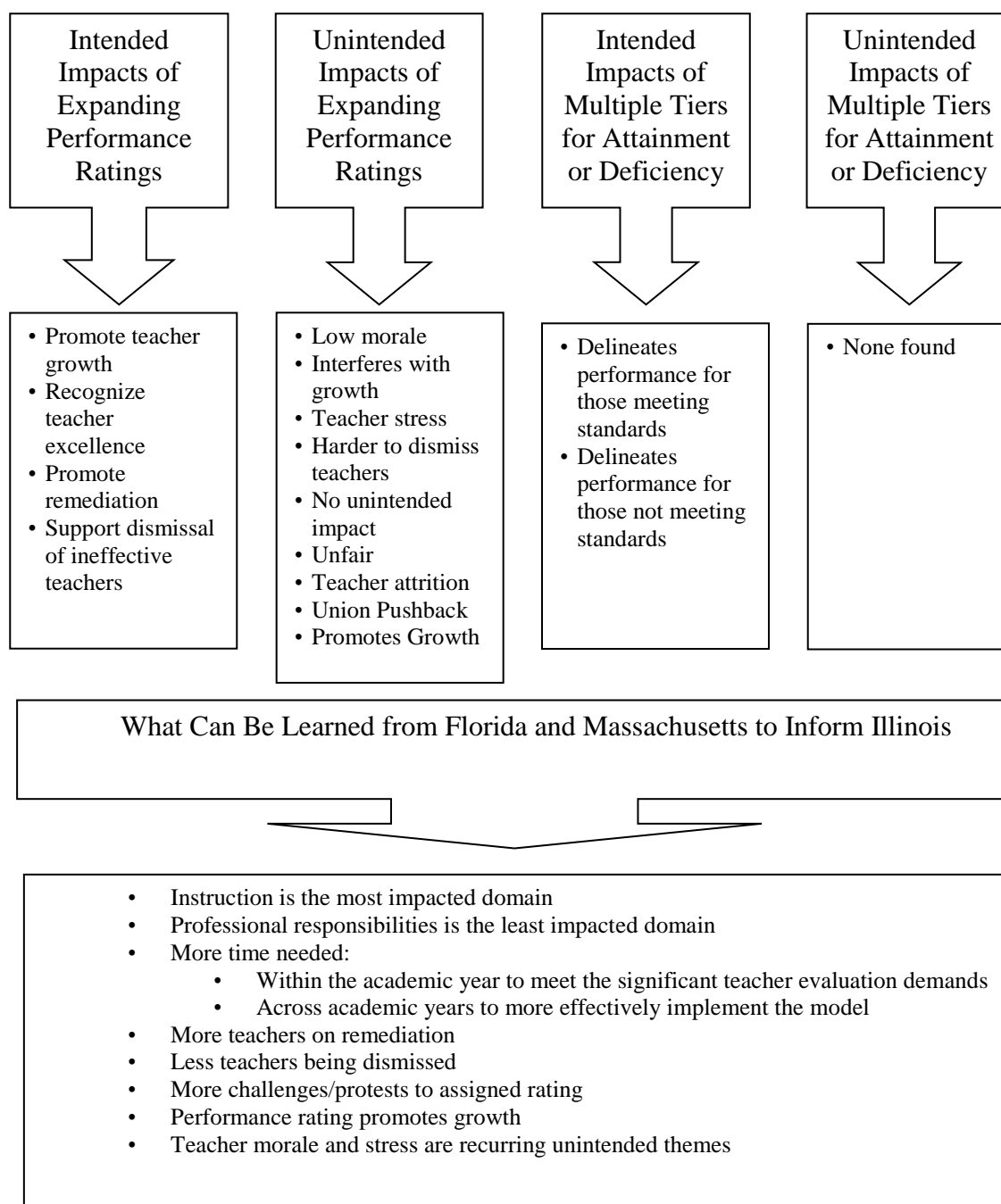
Furthermore, the State of Illinois can learn much from Florida and Massachusetts in light of their recent expansion to a four-tier teacher performance rating evaluation

model. Instruction was conclusively the most impacted domain while the domain of professional responsibilities was clearly the least impacted domain. The study data also revealed a need for more time for principals; both within the academic year to meet the significant teacher evaluation demands and across academic years to more effectively implement the model. More teachers were on remediation while fewer teachers were being dismissed. Principals reported there were more challenges to the assigned teacher evaluation rating given but these principals still asserted the performance rating given did promote growth (see Figure 44).

In applying this work to the field of educational leadership there are at least three practical applications learned from this study. First, the study reinforced the importance of professional development approached with a growth mindset. The study demonstrated that instruction is the most impacted domain in the expanded four-tier rating system so principals need to maximize this impact and believe that all teachers can grow and, in turn, provide the appropriate professional development experiences so that may grow. Second is the importance of promoting a system that has the courage to remediate teachers who need to improve and to dismiss teachers who fail to improve are a determinant to the students they are supposed to serve. A rating system that provides opportunities for remediation and structures to dismiss underperforming teachers is worthless if the principals do not have the courage to use them to promote what is best for the students and the system as a whole. Third, the study emphasized the criticalness of implementing the teacher evaluation system with fidelity so that those that are truly excellent can be recognized for their efforts and growth is promoted for all that are evaluated regardless of performance. Identifying and promoting growth through the

teacher evaluation process is the primary function of the practice and the failure of an evaluator to do so is at the cost to the school system in which they serve.

It is the hope of the researcher that the changes to the teacher evaluation landscape in Illinois and across the nation are pursued with a genuine focus on the promotion of the ideologies of social justice for the sake of the students we serve. If teachers are growing, or being dismissed for failing to be effective, then this should have an alignment and positive correlation with student growth. The findings of this study promote that positive path. It is encouraging to see that the principals recognized that the intended outcomes of the expanded performance evaluation rating systems are to promote teacher growth, recognize excellence, and promote remediation. It is also encouraging to see that these same principals support the dismissal of those teachers who are ineffective and therefore potentially doing harm to our students. Although there are unintended impacts that may be interfering with growth, and having an adverse effect on teachers, the study still found that instruction was positively impacted across all areas studied with teacher growth ultimately being promoted. If the quality of our teachers is improving that means the achievement of our students is also improving which is the ultimate goal of our profession.



*Figure 44. Summary of Study Findings*

APPENDIX A  
CONSENT LETTER

## CONSENT TO PARTICIPATE IN RESEARCH

**Project Title:** *The Perceived Impact of Teacher Performance Ratings on the Teacher Evaluation Process: Voices from the Field*

**Researcher:** *Brian Bullis*

**Faculty Sponsor:** *Dr. Marla Israel*

### **Introduction:**

Dear Principal,

You are being asked to take part in a research study being conducted by Brian Bullis for a dissertation under the supervision of Dr. Marla Israel in the Department of Education at Loyola University of Chicago.

You are being asked to participate because your state has implemented a four-tier teacher evaluation system ahead of the state of Illinois. As a K-12 principal your participation in this study will provide administrators in the state of Illinois the chance to learn from your experiences as they follow a similar path of teacher evaluation.

Please read this form carefully and ask any questions you may have before deciding whether to participate in the study.

### **Purpose:**

The purpose of this study is to explore and measure the principals' perception of the intended and unintended impact of teacher performance ratings on teacher recognition, effectiveness, growth, remediation, and dismissal.

### **Procedures:**

If you agree to be in the study, you will be asked to:

- Complete an online survey that should take approximately 10-15 minutes. The questions will be in a multiple choice format with the exception a few optional short answer questions. The survey will ask you to consider how your current teacher evaluation rating system compares to your past teacher evaluation rating system (if applicable) in regards to teacher recognition, effectiveness, growth, remediation, and dismissal in relation in the four teaching domains of planning and preparation, classroom environment, instruction, and professional responsibilities.

### **Risks/Benefits:**

There are no foreseeable risks involved in participating in this research beyond those experienced in everyday life.

There are no direct benefits to you from participation, but information provided will further inform the following research areas:

- What do principals believe are the intended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- What do principals believe are the unintended impacts of expanding teacher performance ratings to four levels on teacher recognition, effectiveness, growth, remediation, and dismissal?
- What do principals believe are the intended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e. “Excellent,” “Proficient”) or standard deficiency (i.e. “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- What do principals believe are the unintended impacts of providing multiple teacher performance ratings to recognize standard attainment (i.e. “Excellent,” “Proficient”) or standard deficiency (i.e. “Needs Improvement,” “Unsatisfactory”) on teacher recognition, effectiveness, growth, remediation, and dismissal?
- What lessons can be learned to inform educational leaders in Illinois who have recently expanded the number of ratings in their own expanded teacher performance rating system?

**Confidentiality:**

- The survey will not ask for personal contact information and it will not be traceable back to the participant to assure anonymity.
- Survey Monkey® will be used as the instrument for online survey administration and data collection. This is a secure site that provides features to ensure safety and anonymity while administering the surveys and collecting data. Surveys will be treated as private data owned by the user who will use these data for the only for the purpose of this study.

**Voluntary Participation:**

Participation in this study is voluntary. If you do not want to be in this study, you do not have to participate. Even if you decide to participate, you are free not to answer any question or to withdraw from participation at any time without penalty.

**Contacts and Questions:**

If you have questions about this research study, please feel free to contact Brian Bullis at 319-325-0937. You may also contact Dr. Marla Israel, my dissertation director at Loyola University at 312-915-6336 if you have questions or concerns regarding the validity of this study.

If you have questions about your rights as a research participant, you may contact the Loyola University Office of Research Services at (773) 508-2689.

**Statement of Consent:**

You will be asked to consent electronically on the following page. Your electronic consent indicates that you have read the information provided above, have had an opportunity to ask questions, and agree to participate in this research study.

Thank you for your participation.

Sincerely,

Brian Bullis, Doctoral Candidate, Loyola University Chicago

APPENDIX B  
SURVEY INSTRUMENT

### Electronic Consent

You understand that your participation in this research study is voluntary, and you consent to participate in this survey. You understand that you can withdraw your participation at any time. You understand that your responses will remain confidential and anonymous. All data will be stored in a password protected electronic format. To help protect your confidentiality, the survey will not contain information that will personally identify you. The results of this study will be used for scholarly purposes only.

Confidentiality will be maintained to the degree permitted by the technology used. Your participation in this online survey involves risks similar to a person's everyday use of the Internet. When a participant completes this anonymous survey and submits it to the researcher the researcher will be unable to extract anonymous data from the database should the participant choose to withdraw.

By clicking on the "agree" button below I indicate that:

- I have read the above information.
- I voluntarily agree to participate in this study.

By clicking "disagree" I will decline participation and will be exited from this study.

[Check agree or disagree]

### **Survey on the Impact of Teacher Evaluation Ratings**

#### Demographic

- In what state did you evaluate teachers at a public school in the past year?  
*Florida*  
*Massachusetts*  
*I did not evaluate teachers at a public school in either state in the past year*
- What is the highest educational degree you have obtained related to the field of education?  

<i>BA/BS</i>	<i>MA/MS</i>	<i>EdS/EdD/PhD</i>
--------------	--------------	--------------------
- What is your current age?  

<i>20-29</i>	<i>30-39</i>	<i>40-49</i>	<i>50-59</i>	<i>60+</i>
--------------	--------------	--------------	--------------	------------
- How many years of teaching experience do you have?  

<i>0</i>	<i>1-5</i>	<i>6-10</i>	<i>11-15</i>	<i>16+</i>
----------	------------	-------------	--------------	------------
- How many years have you been a principal?  

<i>0-2</i>	<i>3-5</i>	<i>6-10</i>	<i>11-15</i>	<i>16+</i>
------------	------------	-------------	--------------	------------
- What is the approximate number of teachers you evaluate per year?  

<i>0-2</i>	<i>3-5</i>	<i>6-10</i>	<i>11-15</i>	<i>16 or more</i>
------------	------------	-------------	--------------	-------------------
- How many total years have you evaluated teachers using a performance rating system with **four-tiers**?  

<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4 or more</i>
----------	----------	----------	----------	------------------
- How many performance rating tiers did you use to rate teachers **prior** to using your current four-tier model?

0 (if 0 then please bypass any questions asking you to provide feedback on the four-tier model)

1

2

3

4 or more

- For how many years did you evaluate teachers under the rating system you indicated in the previous answer?  

0	1	2	3	4 or more
---	---	---	---	-----------
- Not including the 2013-14 academic year, how many years have you evaluated teachers in your state?  

0-2	3-5	6-10	11-15	16 or more
-----	-----	------	-------	------------

### Overview

- Under your current four-tier teacher rating system, **compared to your prior system**, indicate the extent to which the four levels allow you to:
  - Recognize excellent teachers:  
*Far More Effectively – More Effectively – No Impact – Less Effectively – Far Less Effectively*
  - Identify areas for teacher growth:  
*Far More Effectively – More Effectively – No Impact – Less Effectively – Far Less Effectively*
  - Motivate teachers to grow:  
*Far More Effectively – More Effectively – No Impact – Less Effectively – Far Less Effectively*
  - Identify and recommend teachers for remediation:  
*Far More Effectively – More Effectively – No Impact – Less Effectively – Far Less Effectively*
  - Terminate ineffective teachers:  
*Far More Effectively – More Effectively – No Impact – Less Effectively – Far Less Effectively*
- If it were up to you, how many performance ratings would you prefer to use for teacher evaluation?
  - 0
  - 2 (i.e. *Unsatisfactory, Proficient*)
  - 3 (i.e. *Unsatisfactory, Proficient, Excellent*)
  - 4 (i.e. *Unsatisfactory, Needs Improvement, Proficient, Excellent*)
  - 5 (i.e. *Unsatisfactory, Needs Improvement, Proficient, Accomplished, Excellent*)
  - 6 or more

### Teacher Recognition

- In what domains do you believe your current four-tier teacher rating system, **compared to your prior teacher rating system**, can more effectively recognize teacher excellence: (check all that apply)
  - *Planning and Preparation*
  - *Classroom Environment*
  - *Instruction*
  - *Professional Responsibilities*
  - *None*

- According to your experiences what do you believe are the intended impacts of your state's four-tier rating system in relation to teacher recognition?
- According to your experiences what do you believe have been the unintended impacts of your state's four-tier rating system in relation to teacher recognition?

### Teacher Effectiveness

- In what domains do you believe your current four-tier teacher rating system, **compared to your prior teacher rating system**, can more effectively recognize teacher effectiveness: (check all that apply)
  - *Planning and Preparation*
  - *Classroom Environment*
  - *Instruction*
  - *Professional Responsibilities*
  - *None*
- Do you believe that it is necessary to have a system with more than one tier to recognize teacher attainment of standards (i.e. "Proficient" and "Excellent")?  
*Yes – No ----- Please Explain*
- According to your experiences what do you believe are the intended impacts of your state's four-tier rating system in relation to teacher effectiveness?
- According to your experiences what do you believe have been the unintended impacts of your state's four-tier rating system in relation to teacher effectiveness?

### Teacher Growth

- In what domains do you believe your current four-tier teacher rating system, **compared to your prior teacher rating system**, can more effectively promote teacher growth : (check all that apply)
  - *Planning and Preparation*
  - *Classroom Environment*
  - *Instruction*
  - *Professional Responsibilities*
  - *None*
- How do you feel that the performance rating you assign a given teacher impacts teacher growth?  
*Always Promotes – Usually Promotes – No Impact – Usually Distracts – Always Distracts*
- How do you feel the performance rating promotes the identified areas for growth in the evaluation?  
*Significant Positive Impact – Some Positive Impact – No Impact – Some Negative Impact – Significant Negative Impact*
- How many times per year have teachers challenged or protested the performance rating you have assigned them under the four-tier rating system?  
0                      1                      2                      3                      4 or more
- How many times per year did teachers challenge or protest the performance rating you assigned them under your previous evaluation system?  
0                      1                      2                      3                      4 or more

- According to your experiences what do you believe are the intended impacts of your state's four-tier rating system in relation to teacher growth?
- According to your experiences what do you believe have been the unintended impacts of your state's four-tier rating system in relation to teacher growth?

### Teacher Remediation

- In what domains do you believe your current four-tier teacher rating system, **compared to your prior teacher rating system**, can more effectively identify teachers for remediation: (check all that apply)
  - *Planning and Preparation*
  - *Classroom Environment*
  - *Instruction*
  - *Professional Responsibilities*
  - *None*
- How many teachers per year have you put on remediation under your current four-tier rating system?  
                     0                    1                    2                    3                    4 or more
- How many teachers per year did you put on remediation under your previous evaluation system?  
                     0                    1                    2                    3                    4 or more
- Do you believe that it is necessary to have a system with more than one tier to identify teacher deficiency in meeting standards (i.e. Needs Improvement and Unsatisfactory)?  
                     Yes – No ----- Please Explain
- According to your experiences what do you believe are the intended impacts of your state's four-tier rating system in relation to teacher remediation?
- According to your experiences what do you believe have been the unintended impacts of your state's four-tier rating system in relation to teacher remediation?

### Teacher Dismissal

- In what domains do you believe your current four-tier teacher rating system, **compared to your prior teacher rating system**, can more effectively identify teachers for dismissal: (check all that apply)
  - *Planning and Preparation*
  - *Classroom Environment*
  - *Instruction*
  - *Professional Responsibilities*
  - *None*
- How many teachers have you terminated per year under your current four-tier rating system?  
                     0                    1                    2                    3                    4 or more
- How many teachers have you terminated per year under your previous evaluation system?  
                     0                    1                    2                    3                    4 or more
- According to your experiences what do you believe are the intended impacts of your state's four-tier rating system in relation to teacher dismissal?

- According to your experiences what do you believe have been the unintended impacts of your state's four-tier rating system in relation to teacher dismissal?

Optional: If you would like to further explain any of the answers you have provided in this survey please do so below:

Thank you!

APPENDIX C  
FOLLOW-UP LETTER #1

Dear Principal,

This letter is meant to serve as a follow-up request to participate in an electronic survey.

As a doctoral candidate at Loyola University of Chicago I am conducting research for my dissertation entitled, *The Perceived Impact of Teacher Performance Ratings on the Teacher Evaluation Process: Voices from the Field*. The purpose of this study is to explore and measure the principals' perception of the intended and unintended impacts of teacher performance ratings on teacher recognition, effectiveness, growth, remediation, and dismissal.

If you have already submitted the electronic survey emailed to you one week ago, thank you. If not, please click on the link below to complete the survey. One more reminder will be sent out in approximately one week.

Sincerely,

Brian Bullis

APPENDIX D  
FOLLOW-UP LETTER #2

Dear Principal,

This letter is meant to serve as a final follow-up request to participate in an electronic survey. As a doctoral candidate at Loyola University of Chicago I am conducting research for my dissertation entitled, *The Perceived Impact of Teacher Performance Ratings on the Teacher Evaluation Process: Voices from the Field*. The purpose of this study is to explore and measure the principals' perception of the intended and unintended impacts of teacher performance ratings on teacher recognition, effectiveness, growth, remediation, and dismissal.

If you have already submitted the electronic survey emailed to you two weeks ago, thank you. If not, please click on the link below to complete the survey. The survey will close within the next two days.

Sincerely,

Brian Bullis

## REFERENCES

- Cogan, M. (1973). *Clinical supervision*. Boston, MA: Houghton Mifflin.
- Collins, J. (2005). *Good to great and the social sectors*. Boulder, CO.
- Coleman, E. (1945). The supervisory visit. *Educational Leadership*, 2(4), 164-167.
- Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed method approaches*, 3<sup>rd</sup> ed. Thousand Oaks, CA: Sage Publications, Inc.
- Conley, S., & Glasman, N. (2008). Fear, the school organization, and teacher evaluation. *Educational Policy*, 22, 63-85.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*, 2<sup>nd</sup> ed. Alexandria, VA: ASCD.
- Danielson, C. (2008). *The handbook for enhancing professional practice: Using the framework for teaching in your school*. Alexandria, VA: ASCD.
- Danielson, C. (2011). *The framework for teaching evaluation instrument*. Princeton, NJ: The Danielson Group.
- Danielson, C. (2013). *The framework for teaching evaluation instrument*. Princeton, NJ: The Danielson Group.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston, MA: Henry Holt.
- Ericsson, K., Krampe, R., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Florida Department of Education. (2012, May). *Teacher evaluation systems alignment: The Florida Educator Accomplished Practices (FEAPS), Robert Marzano, & Charlotte Danielson*. Retrieved from the Florida Department of Education website: [https://www.floridaschoolleaders.org/FEAP\\_Alignment\\_Danielson](https://www.floridaschoolleaders.org/FEAP_Alignment_Danielson)

- Florida Department of Education. (2013, January). *Florida's race to the top: A second year progress report on Florida's race to the top*. Retrieved from the Florida Department of Education website:  
[http://www.fldoe.org/arra/pdf/Fla\\_RTTT\\_Final\\_1\\_31.pdf](http://www.fldoe.org/arra/pdf/Fla_RTTT_Final_1_31.pdf)
- Florida Legislative Statutes, K-20 Education Code, Title XLVIII § 1012.34 (2010).
- Florida Legislative Statutes, K-20 Education Code, Title XLVIII § 1012.34 (2011).
- Fullan, M. (2010). *All systems go: The change imperative for whole system reform*. Thousand Oaks, CA: Corwin.
- Gardner, H. (1983). *Multiple intelligences: The theory in practice*. New York: Basic Books.
- Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York: Holt, Rinehart & Winston.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hunter, M. (1980). Six types of supervisory conferences. *Educational Leadership*, 37(5), 408-412.
- Hunter, M. (1982). *Mastery teaching*. El Segundo, CA: TIP Publications.
- Hunter, M. (1984). Knowing, teaching, and supervising. In *Using what we know about teaching* (pp. 169-192). Alexandria, VA: Association for Supervision and Curriculum Development.
- Illinois Pension Code Senate Bill 7, S. Ill. § 097-0008 (2011).
- Illinois State Board of Education. (2012). *Illinois race to the top phase 3: Allocations for participating leas*. Retrieved from  
<http://www.isbe.state.il.us/racetothetop/PDF/phase3-partic-lea-alloc.pdf>
- Illinois State Board of Education. (2013). *Performance Evaluation Advisory Council (PEAC): General faqs about teacher and principal evaluation*. Retrieved from  
<http://www.isbe.state.il.us/peac/html/faqs.htm>
- Illinois State Board of Education. (2010). *The state of Illinois race to the top application for initial funding*. Retrieved from  
<http://www2.ed.gov/programs/racetothetop/phase1-applications/illinois.pdf>

- Illinois State Board of Education. (2010). *The state of Illinois race to the top application for phase 2 funding*. Retrieved from <http://www2.ed.gov/programs/racetothetop/phase2-applications/illinois.pdf>
- Illinois State Board of Education. (2011). *The state of Illinois race to the top application for phase 3 funding*. Retrieved from <http://www2.ed.gov/programs/racetothetop/phase3-applications/illinois-2.pdf>
- Israel, M., & Kersten, T. (2007). Targeted staff development on writing instruction: building multiple theories to provide a comprehensive practical model. *School Leadership Review*, 2(3), 34-61.
- Jackson, R. (2008). *The instructional leader's guide to strategic conversations with teachers: How to provide customized feedback to teachers that actually helps them get better at teaching*. Washington, DC: Mindsteps, Inc.
- Johnson, S.M., & Donaldson, M.L. (2006). The effects of collective bargaining on teacher quality. In J. Hannaway, & A.J. Rotherman (Eds.), *Collective bargaining in education: Negotiating change in today's schools* (pp. 111-140). Cambridge, MA: Harvard Education Press.
- Keigher, A. (2010). *Teacher attrition and mobility: Results from the 2008-09 teacher follow-up survey* (NCES 2010-353). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>.
- Kersten, T., & Israel, M. (2005). Teacher evaluation: principals' insights and suggestions for improvement. *Planning and Changing*, 36, 47-67.
- Kimball, S., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Education Administration Quarterly*, 45, 34-70.
- Kyte, G. (1930). *How to supervise*. Boston, MA: Houghton-Mifflin.
- Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. New York: The Wallace Foundation.
- Loyola University Institutional Review Board (2013). *Informed consent*. Retrieved from <http://www.luc.edu/irb/irbinformedconsent.shtml>
- Loyola University Institutional Review Board (2013). *Online survey research*. Retrieved from <http://www.luc.edu/irb/irbonlinesurveys2.shtml>

- Marshall, C., & Oliva, M. (2010). *Leadership for social justice: Making revolutions in education*, 2<sup>nd</sup> ed. Boston, MA: Pearson Education, Inc.
- Marshall, C., & Rossman, G. (2011). *Designing qualitative research*. Thousand Oaks, CA: Sage Publications, Inc.
- Massachusetts Department of Elementary and Secondary Education. (1995). *Advisory on school governance*. Retrieved from <http://www.doe.mass.edu/lawsregs/advisory/cm1115gov.html#per>
- Massachusetts Department of Elementary and Secondary Education. (2011). *Building a breakthrough framework for educator evaluation in the commonwealth*. Retrieved from <http://www.doe.mass.edu/eval/breakthroughframework.pdf>
- Massachusetts Department of Elementary and Secondary Education. (2012). *Educator evaluation: frequently asked questions*. Retrieved from <http://www.doe.mass.edu/eval/faq.html?section=all>
- Massachusetts Department of Elementary and Secondary Education. (2011). *Overview of the new Massachusetts educator evaluation framework*. Retrieved from <http://www.doe.mass.edu/eval/101511Overview.pdf>
- Massachusetts Department of Elementary and Secondary Education. (2011). *Status of the Massachusetts educator workforce*. Retrieved from <http://www.doe.mass.edu/research/reports/1211edworkforce.pdf>
- Massachusetts Department of Elementary and Secondary Education, Education Laws and Regulations, 603 CMR § 35.08 (2011).
- Massachusetts Teacher Association. (2011). *Reinventing educator evaluation: Connecting professional practice with student learning*. Retrieved from the Massachusetts Teacher Association website: <http://www.massteacher.org/~media/Files/PDFs/CEPP/evalreport.pdf>
- Marzano, R., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results*. Alexandria, VA: ASCD.
- Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.
- May, H., & Supovitz, J.A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, 47, 332-352.
- McQuinn, P. (2012). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems*. Retrieved

from the Center for American Progress website: [http://www.americanprogress.org/wp-content/uploads/2012/11/McGuinn\\_TheStateofEvaluation-1.pdf](http://www.americanprogress.org/wp-content/uploads/2012/11/McGuinn_TheStateofEvaluation-1.pdf)

Merriam, S.B. (2009). *Qualitative research: A guide to design and implementation*, 5<sup>th</sup> ed. San Francisco: Jossey-Bass.

Metlife (2009). *The metlife survey of the American teacher. Collaborating for student success, part 3: Teaching as a career*. Retrieved from [https://www.metlife.com/assets/cao/contributions/foundation/american-teacher/MetLife\\_Teacher\\_Survey\\_2009\\_Part\\_3.pdf](https://www.metlife.com/assets/cao/contributions/foundation/american-teacher/MetLife_Teacher_Survey_2009_Part_3.pdf)

Murphy, J., Hallinger, P., & Heck, R.H. (2013). Leading via teacher evaluation: the case of the missing clothes? *Educational Researcher*, 42, 349-354.

National Comprehensive Center for Teacher Quality. (2012). *Databases on state teacher and principal evaluation policies (STEP database and SPEP database)*. Retrieved from <http://resource.tqsource.org/stateevalldb/>

National Council on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Retrieved from the National Council on Teacher Quality website: [www.nctq.org/p/publications/docs/nctq\\_stateOfTheStates.pdf](http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf)

Nolan, J., & Hoover, L.A. (2005). *Teacher supervision and evaluation: Theory into practice*. Hoboken, NJ: John Wiley & Sons, Inc.

Palmer, P. (2007). *The courage to teach: Exploring the inner landscape of a teacher's life*. San Francisco, CA: Jossey-Bass.

Performance Evaluation Advisory Council. (2011, June). *Principal and teacher evaluation in Illinois: Past, present, and future*. Retrieved from Illinois State Board of Education: [www.isbe.state.il.us/PEAC/ppt/peac-pera\\_slides.pptx](http://www.isbe.state.il.us/PEAC/ppt/peac-pera_slides.pptx)

Performance Evaluation and Reform Act of 2010, S. Ill. § 096-0861 (2010).

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin*, 88, 60-79.

Pink, D. (2009). *Drive: The surprising truth about what motivates us*. New York: Penguin Group.

Reeves, D. (2004). *Assessing educational leaders*. Thousand Oaks, CA: Corwin Press.

Sartain, L., Stoelinga, S.R., & Krone, E. (2010). *Rethinking teacher evaluation: Findings from the first year of the excellence in teaching project in Chicago public schools*. Retrieved from University of Chicago, Consortium on Chicago School Research

website: <http://ccsr.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Final.pdf>

- Schmoker, M. (2011). *Focus: Elevating essentials to radically improve student learning*. Alexandria, VA: ASCD.
- Schwandt, T. (2007). *The Sage dictionary of qualitative inquiry*, 3<sup>rd</sup> ed. Thousand Oaks: Sage Publications, Inc.
- Scott, S. (2004). *Fierce conversations: Achieving success at work and in life, one conversation at a time*. New York: The Berkley Publishing Group.
- Sergiovanni, T. (1992). *Moral leadership: Getting to the heart of school improvement*. San Francisco, CA: Jossey-Bass, Inc.
- Sergiovanni, T., & Starratt, R. (2007). *Supervision: A redefinition*, 8<sup>th</sup> ed. New York: McGraw-Hill.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.
- Starratt, R. (2012). *Cultivating an ethical school*. New York: Routledge.
- Stronge, J., Gareis, C., & Little, C. (2006). *Teacher pay and teacher quality: Attracting, developing, and retaining the best teachers*. Thousand Oaks, CA: Corwin Press.
- Survey Monkey. (2012). *Privacy policy*. Retrieved from <http://www.surveymonkey.com/mp/policy/privacy-policy/>
- Taylor, F. (1911). *The principles of scientific management*. Reprinted 2007, Sioux Falls, SD: NuVision.
- The KIPP Foundation. (2013). *The KIPP foundation*. Retrieved from The KIPP Foundation website: <http://www.kipp.org/about-kipp/the-kipp-foundation>
- The New Teacher Project. (2010). *Teacher evaluation 2.0*. Retrieved from The New Teacher Project website: <http://tntp.org/assets/documents/Teacher-Evaluation-Oct10F.pdf>
- The Wallace Foundation. (2011). *The Wallace Foundation launches major 'principal pipeline' initiative to help school districts build corps of effective school leaders*. Retrieved from The Wallace Foundation website: <http://www.wallacefoundation.org/view-latest-news/PressRelease/Pages/The-Wallace-Foundation-Launches-Major-Principal-Pipeline-Initiative-to-Help-School-Districts-Build-Corps.aspx>

- Thomas, E., & Wingert, P. (2010, March). Why we must fire bad teachers. *Newsweek*. Retrieved from [www.newsweek.com/id/234590](http://www.newsweek.com/id/234590)
- Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.
- Tucker, P., & Stronge, J. (2005). *Linking teacher evaluation and student learning*. Alexandria: ASCD.
- U.S. Department of Education. (2012). *Race to the top: Massachusetts report, year one: School year 2010-2011*. Retrieved from [www2.ed.gov/programs/racetothetop/performance/ Massachusetts-year-1.pdf](http://www2.ed.gov/programs/racetothetop/performance/Massachusetts-year-1.pdf)
- U.S. Department of Education. (2009). *Race to the top program executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education, Office of the Deputy Secretary. (2011). *Race to the top: Phase 3 award letter, Illinois*. Retrieved from <http://www2.ed.gov/programs/racetothetop/phase3-awards/illinois.pdf>
- Wagner, C.A. (1921). *Common sense in school supervision*, Milwaukee, WI: The Bruce Publishing Company.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Retrieved from The New Teacher Project website: [http://widgeteffect.org/downloads/TheWidget Effect.pdf](http://widgeteffect.org/downloads/TheWidgetEffect.pdf)
- Whitehead, M. (1952). Teachers look at supervision. *Educational Leadership*, 10(2), 1011-1106.
- Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.
- Zepeda, S.J. (2007). *The principal as instructional leader: A handbook for supervisors*, 2<sup>nd</sup> ed. Larchmont, NY: Eye on Education, Inc.

## VITA

Brian Bullis was born and raised in Waterloo, Iowa. Before attending Loyola University Chicago, he attended the University of Iowa where he earned a Bachelor of Arts in History in 2002 and a Masters of Arts in Educational Administration in 2007.

Bullis has taught World History, U.S. History, Economics, and Psychology at West Liberty High School in West Liberty, Iowa. He also served as a Dean of Students, Activities Director, and Assistant Principal at West Liberty High School and West Liberty Middle School. Bullis also served as Assistant Principal at Alan B. Shepard Middle School in Deerfield, Illinois. Currently, Bullis is a principal at Caruso Middle School in Deerfield, Illinois.

## DISSERTATION COMMITTEE

The Dissertation submitted by Brian Bullis has been read and approved by the following committee:

Marla Israel, Ed.D., Director  
Associate Professor, School of Education  
Loyola University Chicago

Janis Fine, Ph.D.  
Associate Professor, School of Education  
Loyola University Chicago

Susan Sostak, Ed.D.  
Clinical Assistant Professor, School of Education  
Loyola University Chicago