A comparison of teacher evaluation systems in New York State and Finland - are there alternatives to APPR?

Lindsay Tresansky
University at Albany, State University of New York, ltresansky@albany.edu

A COMPARISON OF TEACHER EVALUATION SYSTEMS

IN NEW YORK STATE AND FINLAND —

ARE THERE ALTERNATIVES TO APPR?

by

Lindsay M. Tresansky

A Dissertation

Submitted to the University at Albany, State University of New York

In Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

School of Education

Department of Educational Policy and Leadership

2023

ABSTRACT

The Annual Professional Performance Review (APPR) system in New York State (NYS) has been called into question by educators since its adoption nearly 10 years ago, yet it remains the mandated evaluation system in NYS schools today. Much of the concern has been over changes such as assigning teachers final evaluation scores, as well as for the added emphasis of students' state assessment results on teachers' individual final evaluation scores. In fact, the US Secretary of Education continues to promote the importance of administering high-stakes standardized tests, in spite of educator feedback expressing concerns on the impacts of the COVID-19 pandemic on student performance (Strauss, 2021). The purpose of this study was to understand potential alternatives to the APPR teacher evaluation system as practiced in New York State and elsewhere in the United States, such as the quality assurance model in Finland.

This study sought to understand the range of possibilities for teacher evaluation. Finland is regarded as having an excellent education system (Sahlberg, 2011). Finland is known for its policies and practices that are deeply rooted in trust in teachers, collaboration between professionals, and the wellbeing of all stakeholders. Examining the perceptions of NYS and Finnish teachers and supervisors through interviews in this qualitative study was insightful, due to the subjects' proximity to the teaching and learning processes. A sample of eleven educators, five from Finland and six from NYS, each with varied ranges of experience and content certifications were interviewed. The study also investigated feedback and the aspects of the teaching profession that motivated teachers to improve. In examining evaluation from a comparative perspective, this qualitative research study may contribute to laying the groundwork for making policy recommendations for the APPR teacher evaluation system as practiced in NYS and elsewhere in the United States.

ACKNOWLEDGEMENTS

I now realize that a completed dissertation is very much a reflection of the vast and caring support network surrounding and uplifting the researcher. I am fortunate to have been surrounded by many caring and giving individuals who have supported me along my journey. First, I would like to thank my dissertation chair, Dr. Heinz-Dieter Meyer. I am grateful for your presence in the EPL program, and for your belief in me as a researcher and practitioner. Your commitment to research and making the world wiser through your efforts in the field of education is inspiring. We all benefit from your strength and good ideas.

Thank you also to my fantastic dissertation committee members, namely Dr. Kathie Spring, Dr. Jerusalem Rivera-Wilson, and Dr. Eija Rougle. You have challenged me, broadened my thinking, and extended the boundaries of my world. I am grateful for your investment in me. Also, a special thank you to Dr. Eija Rougle and Dr. Katri Karasma for all of your assistance and support in helping me plan my trip to Helsinki, Finland. Additionally, a big thank you to my fellow doctoral candidates, Fred Engelhardt and Zak Snyder. Your feedback and supportive conversations have made all the difference in helping me cross the finish line. I will always look back on our experiences in this program together with fondness.

I am also very fortunate to have a support network that includes many professional colleagues in K-12 education. Thank you very much to Dr. David Perry, Timothy Backus, Christopher Robilotti, Jacqlene McAllister, and the South Colonie Board of Education for your support of me as I completed my dissertation and Ph.D. Your encouragement and positivity throughout this journey have meant the world to me. I am thankful to work for a school district that is committed to the professional growth and development of its administration and staff.

My professional support network also extends outside the bounds of South Colonie. I

would like to thank a few people who have been especially influential to me along the way. First, a big thank you to Kathleen Skeals for always supporting my career and professional aspirations from the very beginning. Your presence in my life has given me so much strength. Second, thank you to Dr. Elizabeth Wood for giving me the necessary "push" to pursue this degree at this time. It has been anything but easy completing this degree while becoming a building principal and raising two young children, and I could not have done it without someone believing that I had the capacity to do so. Finally, a big thank you to Dr. David Hurst for your encouragement to apply for admission to this excellent program, and for all of your unwavering support and advice in navigating the nuances of this program along the way.

Last, and most importantly, I am so incredibly grateful for my loving family. Thank you to my husband, Tom, and my children, Lucas and Addy, for your love, patience, and understanding as I took on this degree, which has not been without impact on our family. Thank you for all of the quiet time you gave me to write, especially when taking the doctoral exam and also when finishing the home stretch of the dissertation. There is no way I could have done this without an immense amount of patience, flexibility, and understanding on your part. Your belief and support in me have been overwhelming and humbling. I am greatly looking forward to having more time to spend with you as a result of completing this degree.

Thank you also to my parents, Roland and Janine, for all of your love, support, and encouragement. My wellbeing has always been your number one priority, especially when taking on something new, and I am grateful for all you always do to patiently listen to me and ask me what I need to meet with success. However large my aspirations have been along the way, you always encouraged me. I am so grateful for the confidence you have instilled in me throughout my life, and I only hope I can one day instill that same sense of confidence in my own children.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Introduction

The Annual Professional Performance Review (APPR) is the current standardized teacher evaluation system used by school districts in New York State. It was launched in 2013 as part of the federal Race to the Top initiative. State policy outlines the following formula to determine a teacher's annual evaluation score: 60% state-approved rubrics, 20-25% state assessment results, and 15-20% local assessment results (New York State Legislature, 2010). The 60% rubrics portion varies by school district; however, it typically consists of teacher observations, professional responsibilities, and professional development (New York State Education Department, n.d.). In addition to the 20-25% state assessments portion, state assessment results may also apply to the 15-20% local assessment piece, depending on the course load of the individual teacher. This could result in 40% of the evaluation of a teacher being dependent upon the performance of his or her students on high-stakes standardized tests. Since its adoption nearly 10 years ago, policy makers have met with skepticism over APPR, particularly for the emphasis of students' state assessment results on teachers' final scores. Assigning teachers final evaluation scores is another APPR practice that has been questioned. Many uncertainties remain: Does the current APPR system measure quality teaching? Does APPR support teacher growth and motivate educators to develop professionally? Are there alternative systems to APPR?

Seeking to understand the range of possibilities for assessing teacher quality is what led to me to study teacher evaluation systems around the world, and in particular, Finland. While many options exist that are worthy of study and comparison to New York State's APPR system, Finland has been recognized as one of the most well-developed education systems in the world (Schwab & Zahidi, 2020). In addition, Finland is regarded as one of the world's most literate societies and

is heralded for its international success on the PISA exam (Sahlberg, 2011). Another reason Finland makes for a compelling comparison is that its policies in education are guided by a professional accountability model, as opposed to the market accountability model that guides education policy in the United States (Williams & Engle, 2013). Philosophically, these are very different, and at times opposing accountability approaches. Finland is known for its policies and practices that are deeply rooted in trust in teachers, collaboration between professionals, and the wellbeing of all stakeholders (Kelly et al., 2018).

Kelly et al. (2018) explains that Finnish education policies are written such that they address inequality before students' formal schooling even begins. For example, Finland has a strong early childhood education system for children up to age seven that is heavily subsidized. Finland also has a universal health care system. These investments in the wellbeing of children and society at large have been shown to pay dividends down the road in Finland, where the childhood poverty rate is only 3.6%. This stands in contrast to the United States where the childhood poverty rate is 20.2% (Kelly et al., 2018).

In Finland there is no formalized system or national policy for teacher evaluation. Instead, according to Hammerness et al. (2017), the evaluation of teachers in Finland, sometimes referred to as quality assurance, is focused on professional development at the individual level rather than on student performance on standardized test scores. It is not common practice for Finnish administrators to use a rubric for evaluating teachers. Instead, feedback to teachers is provided face-to-face and is often in the form of informal two-way dialogue with the school leader. During such exchanges, the administrator and the teacher come to agree upon the content. Some of the feedback and discussion focuses on key features of teaching, such as personal performance, versatility, initiative, and ability to cooperate. Hammerness et al. (2017) discusses that part of the

success of the teacher evaluation system in Finland is that school leaders engage in "management by walking around" in their school buildings, and are keenly aware of the work taking place in classrooms between professionals and children. This model is possible because of the closely-networked school communities that are cultivated in Finland (Hammerness et al., 2017).

Other conditions that contribute to Finnish school culture are that teachers are looked at with high social prestige, and are given significant professional autonomy (Hammerness et al., 2017). Furthermore, labor agreements for teachers are negotiated collectively, systems exist that contribute to the development and preparation of teachers, and teachers make a reasonable salary that is commensurate with other professions. In general, teaching is viewed as a service to society and the public good in Finland (Tarhan et al., 2019). Finland is a unique place with a unique history, and is both happy and healthy but also a "cultural lone wolf," in that it tends to find its own way to do things rather than follow the paths of others (Sahlberg, 2021). Although the United States is culturally and geographically distinct from Finland, there are still many insightful points that can be learned from Finland's experience with educational change.

Teacher Evaluations in New York State Before APPR

Prior to the implementation of the APPR teacher evaluation system in New York State (NYS) in 2013, teacher evaluations looked quite different. There was no state-level standardized evaluation system in place. Each individual school district had their own local method of evaluating teachers, and these methods widely varied. For example, some school districts used rubrics or templates, and some did not. If there was a rubric or template, this was sometimes agreed upon between the school district and the teachers' collective bargaining unit. Feedback was often qualitative in nature and it was not a common practice to assign teachers with a final score other than perhaps an indication that the teacher's performance was satisfactory (Danielson, 2011).

Because teacher evaluations were something governed by each individual school district as a local process, their structure and content was determined by the unique priorities of the teacher's supervisor and/or district leadership team. For example, student achievement may have been one area of consideration in a teacher's overall annual evaluation, but it may have sat alongside several other categories in which supervisors provided the teacher with feedback, depending on the school district. The type of feedback the teacher may have been given on student achievement was potentially qualitative (e.g., teacher's student exceeded expectations on state assessments) or quantitative (e.g., 88% of students scored a level 3 or higher on the state ELA exam).

In the absence of a state standardized teacher evaluation system, local evaluation systems prior to APPR focused more on the development of teachers than on assessment results. While each evaluation process and instrument varied by district, some general commonalities were that strengths and weaknesses were considered, and in some cases, goals were developed or growth plans were written for the following school year. Many areas could have been considered for evaluation, such as content knowledge, classroom techniques and procedures, student achievement, classroom management, teacher/student relations, teacher/colleague relations, teacher/administrator relations, teacher/family relations, educational professionalism, or attendance. In some cases, the supervisor may have provided written feedback to evaluate the teacher's performance in each area. While student achievement was a consideration for teacher evaluation in some school districts, the practice of providing quantitative feedback (such as the use of student test scores as a means of measuring teacher quality) was not commonplace in teacher evaluation systems prior to APPR. In some districts, comparisons were made between the percentages of a teacher's own students who passed (65%) or achieved mastery (85%) on local

final exams or NYS Regents exams, with the departmental averages for each course, but this was not standardized. The number of course failures may have also been considered.

The Benefits of APPR at the Time of Implementation

Leading up to the federal Race to the Top initiative, there was growing concern amongst reformers in the field of education that teachers were not being provided with enough feedback. In fact, Alderman and Chuong (2014) made the case that local teacher evaluation systems pre-APPR assigned teachers with a score of either satisfactory or unsatisfactory, and described this as a 'black and white' way of looking at evaluations. Instituting a system that was more layered and provided teachers with more feedback was a key rationale leading up to APPR implementation.

Other discussion points amongst reformers at the time of APPR implementation were teacher professional development, policy making, high-stakes teacher evaluation, teacher selection, and teacher certification/qualification requirements. Bryk et al. (2012) completed one study where after gathering the different perspectives around a roundtable of stakeholders, several recommendations were generated to address the concerns. The researchers called for methods of rewarding effective teaching, using data and metrics to improve how we select, develop, celebrate, and retain teachers, developing practical measurement tools, and detailing what teachers do in practical terms. The authors referred to this as a "science of improvement" (Bryk et al., 2012, p. 103). The researchers claimed that we lack a common language, a shared framework for measuring progress toward higher standards, and that we have weak capacities to learn collectively from effective teachers. These recommendations for improved systems were all cited as a means of closing the achievement gap and also as a means to learn from one another and improving systems through collaboration (Bryk et al., 2012)

In contrast to many local teacher evaluation systems that had been in place prior to APPR,

the "new" APPR evaluation system was perceived by many to be vastly different, most notably due to the shift in emphasis on students' results on high-stakes standardized test scores to assess the effectiveness of teachers, as well as the increased use of measurement techniques and evaluating teachers quantitatively.

Baker et al. (2010) explains that any sound evaluation will necessarily involve a balancing of many factors to provide a more accurate view of what teachers do in the classroom and how that contributes to student learning. In teacher evaluation systems prior to APPR, evaluation feedback in the area of student success was only one consideration of many in a teacher's evaluation. Now under the APPR system, student achievement data on state standardized exams can determine up to 40% of a teacher's final score. Furthermore, an additional stipulation of the APPR policy allows for the 40% weight assigned to test scores to be used as a means to remove "ineffective" teachers from the profession through an expedited 3020A hearing process (New York State Legislature, 2010).

Research Problem

It is contested that the current APPR teacher evaluation system in New York State captures what makes an effective teacher. In addition, because the APPR system is measurement-driven rather than growth-driven, it is contested that it motivates teachers to improve their craft.

Significance of the Study

There are reasons why this study is of interest to the field. For NYS teachers who have been evaluated using both pre-APPR and APPR practices during their careers, there are differences in each of these experiences. In addition, for NYS supervisors who have evaluated teaching staff using both pre-APPR and APPR practices, there are differences in these experiences as well. The noticeable shift away from local school district-driven criteria to a marked emphasis on state

standardized test results, as well as to measurement-driven methods to assess teacher quality have been the biggest changes APPR has brought. Now that APPR is the standardized teacher evaluation system across New York State, the effectiveness of this system at assessing teacher quality is contested by both teachers and supervisors. Additionally, the effectiveness of APPR to motivate teachers to grow and develop in their profession is called into question.

This study is of interest to the field of education because there are important differences that exist in the experiences of teachers who are evaluated using different evaluation systems, i.e., systems in New York State versus Finland. There are also differences in the experiences of supervisors who are evaluating teachers using different evaluation systems. It is important to understand each of these experiences as part of understanding how evaluation systems impact the field of education, particularly the lives and careers of teachers and supervisors. It is also important to consider the impact that different teacher evaluation systems and practices have on motivation and professional growth and development.

Research Purpose

This qualitative study seeks to understand potential alternatives to APPR, such as the case of Finland. Additionally, this study seeks to understand different types of supervisor feedback and their effect on teacher motivation.

Research Questions

Research Question #1: How do teacher evaluation systems in New York State compare with those in Finland at both the structural and cultural level?

Research Question #2: How do teachers in New York State and Finland view the contribution of teacher evaluations to their motivation and professional growth?

Research Question #3: What are the policy implications resulting from the above

comparison?

Supplemental Research Question #1: How satisfied are teachers and supervisors with teacher evaluation practices in Finland versus New York State?

Supplemental Research Question #2: To what degree do teacher evaluation practices contribute to professional growth in Finland versus New York State?

Definition of Terms

To better understand the review of the literature for this study, the following key definitions are provided:

APPR: Annual Professional Performance Review, the current teacher evaluation system in New York State that was launched in 2013 (New York State Legislature, 2010)

Pre-APPR: Before APPR implementation

Phronesis: Practical knowledge, wisdom, judgment (Meyer, 2016)

Developmental Functions: Developmental tasks include designing activities or lessons, answering content questions, modeling or team-teaching lessons, and facilitating professional development (Gigante & Firestone, 2007). Developmental tasks increase human resources as the teacher leaders contribute to the development of their colleagues' instructional knowledge and skills (Gigante & Firestone, 2007).

Support Functions: Support functions include managing materials or preparing laboratories, building confidence or generating enthusiasm, and piloting curriculum (Gigante & Firestone, 2007). They set the stage for successful teaching by facilitating teachers' work in the short run, but do not increase their human resources in the long run (Gigante & Firestone, 2007).

Comparison of NYS and Finland – Demographics, Geography, Culture

The governance, living conditions, size, and geography of New York State and Finland

vary drastically. In addition, the demographics vary greatly as well. The United States, including New York, has a very socioeconomically and culturally diverse population, whereas the population in Finland is more homogenous. There are also significant cultural differences between Finland and New York State, which will be discussed in a review of the literature as well as in the discussion of the study findings. While these cultural differences would make it nearly impossible to copy the Finnish system in the United States (and New York State), there are still aspects of the Finnish system that could potentially be implemented to make positive change. Some of these potential aspects include increased teacher collaboration, professional development, and teacher autonomy, and these will be explored in this study (Kelly et al., 2018).

To illustrate the geographical differences between the United States and Finland, below is a map of the world. The United States is highlighted in orange, and Finland is highlighted in green (*Finland-United States Relations – Wikipedia*, n.d.).

Figure 1.1: Map of the World
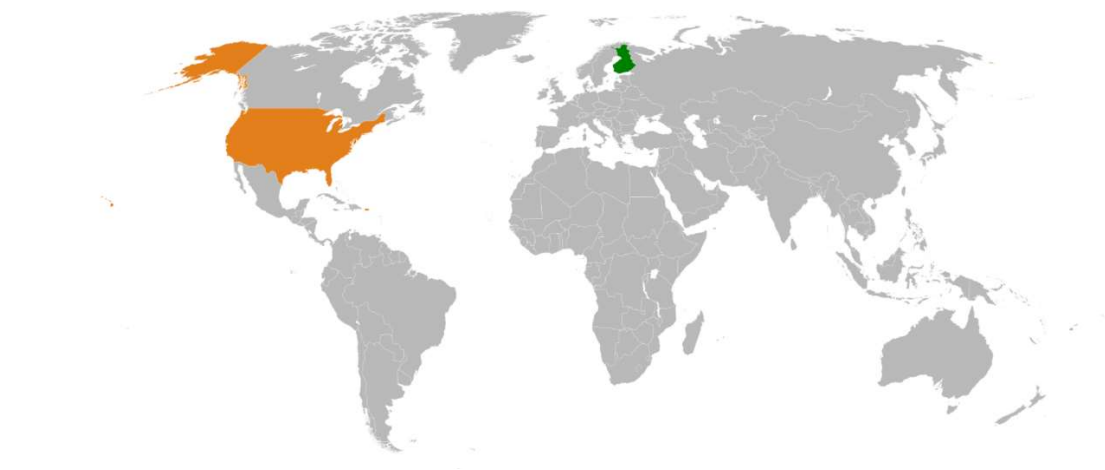
In addition, below are two maps of the United States where an outline of Finland is positioned on top of the map of the United States to illustrate a size comparison of the two countries. The map on the left is helpful for comparing the size of Finland with the size of New

York State (*Size of United States compared to Finland*, n.d.). The map on the right is helpful for comparing the size of Finland with the size of the United States (*Axe FX III | Page 80 | The Gear Page*, n.d.).

Figure 1.2: Map of the Finland in Relation to New York State

Figure 1.3: Map of Finland in Relation to the United States
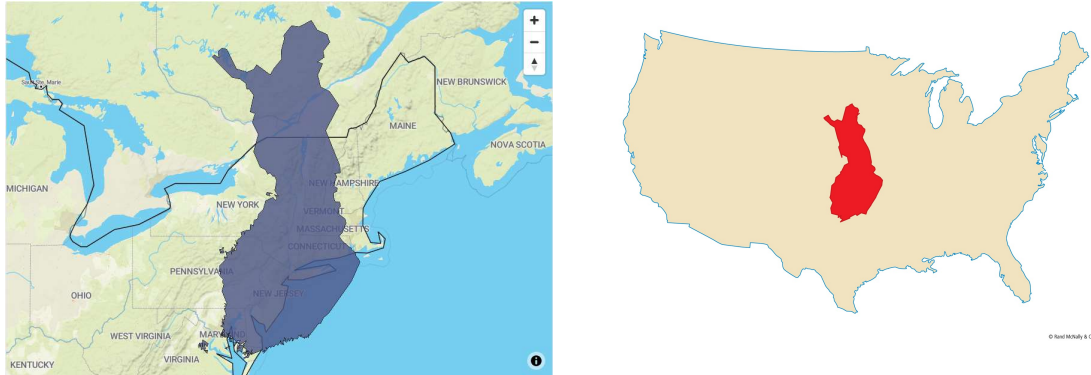
Summary

This chapter defined the research problem and offered reasons why it should be investigated. This chapter also explained the purpose of this study and research questions to be investigated. The next chapter contains a thorough and detailed review of the literature. This will provide the necessary perspective and background for this research study.

CHAPTER 2: LITERATURE REVIEW

Introduction

In this chapter, I will provide a comprehensive review of the relevant literature surrounding APPR in NYS and teacher evaluation practices in Finland. An important gap in the existing literature is that very few studies have examined evaluation processes from a comparative perspective. In addition, little is known about the specific feedback that Finnish supervisors provide to teachers, and how this contributes to development. The three major themes organizing this chapter are, 1. "What is APPR? Does it work? What are the problems with it"; 2. "An alternative to APPR: How does Finland assess teacher quality?"; and 3. "Can New York State Learn from Finland?" The measurement of teacher quality, unintended consequences of APPR, perceptions of teachers, the case of Finland, and evaluation practices that motivate teachers to grow and improve are discussed as well. This exploration and analysis of past research provides me with the necessary background and perspective for conducting my study. The research questions guiding the study are listed below.

Research Question #1: How do teacher evaluation systems in New York State compare with those in Finland at both the structural and cultural level?

Research Question #2: How do teachers in New York State and Finland view the contribution of teacher evaluations to their motivation and professional growth?

Research Question #3: What are the policy implications resulting from the above comparison?

Supplemental Research Question #1: How satisfied are teachers and supervisors with teacher evaluation practices in Finland versus New York State?

Supplemental Research Question #2: To what degree do teacher evaluation practices contribute to professional growth in Finland versus New York State?

Search Strategy

The search strategy for this study started with developing a literature review components concept map and outline. This preparatory work guided the keywords used in search databases. Keywords included, but were not limited to APPR, teacher evaluation, teacher growth, teacher motivation, Finland, continuous improvement, trust, and assessment. I primarily accessed the University at Albany Library's website search tool which generated electronic and print material results. I also utilized Google Scholar and JSTOR Labs Text Analyzer to search for information. My sources of information include peer-reviewed journal articles, books, policy briefs, newspaper articles, blog postings, and press releases from government agencies. Other than those necessary to provide key context or a methodological framework, the majority of my sources have been published within the past twenty years.

What is APPR? Does it work? What are the problems with it?

Theoretical Framework

According to Eisner (1984), teaching should be publicly acknowledged as both an art and a craft, rather than a science. "Teachers are more like orchestra conductors than technicians. They need rules of thumb and educational imagination, not scientific prescriptions" (Eisner, 1984, p. 2). Treating teaching as a science portrays it as something that can be made more efficient and routinized through scientific techniques such as measurement. This conceptualization is also appealing to taxpayers, as it is pleasing to envision schools as capable of producing the products they are expecting. The scientific conceptualization is also embraced by school administrators at times, as it reduces their own susceptibility to public criticism.

Eisner (1984) goes on to say that one of the consequences of portraying teaching as a science is that, "what is educationally significant but difficult to measure or observe is replaced with what is insignificant but comparatively easy to measure or observe" (p. 8). This still holds true with APPR today. Critics of APPR are quick to cite that the evaluation instruments are far too focused on students' performance on high-stakes standardized tests, which are notably easy to measure (Darling-Hammond et al., 2012). Eisner (1984), Labaree (2000), Darling-Hammond et al. (2012), and others purport that the intangibles of effective teaching such as the ability to read the room and take appropriate situational action in order to best advance students' learning forward are significant criteria in which to provide feedback to teachers, yet are practically impossible to quantify and are subsequently often omitted from evaluation instruments such as APPR.

Eisner (1984) makes the case that conceptualizing teaching as both art and craft addresses the "large space between the ideas that science can provide and the kinds of decisions and actions a teacher must take" (p. 9). It means to acknowledge that there will never be a science of teaching that is so prescriptive as to make teaching routine. It also recognizes the classroom as a dynamic space where the teacher must invent or conjure a set of moves to meet the needs of his or her students in the moment and move their learning forward. Applying this to APPR, instead of seeking scientific solutions to educational problems, we should instead view schools as fluid, professional communities and focus policy efforts on creating conditions for teachers to continuously grow as professionals. Nurturing such conditions and creating opportunities for teachers to feel they have used their talents to positively impact their students' growth and development has been shown to be intrinsically motivating to teachers and exceeds whatever motivation it is that sabbaticals or vacations can provide (Eisner, 1984).

Meyer (2016) offers an extension of Eisner's (1984) work on the topic of the limits of

measurement. Meyer's (2016) premise is that the classroom is a domain of practical knowledge, or phronesis, where quality is best appraised by experienced practitioners. This is because practitioners possess the requisite context-sensitive judgment to provide feedback that is meaningful, appropriate, and free from extraneous interest. Meyer (2016) perceptively points out the potential conflicts of interest of those in power who purport to cure educational ills, and how framing education as a science and establishing ranking systems to evaluate educational practices, such as high-stakes standardized assessments, opens the door to extraneous interest in education.

Meyer (2016) explains that practical knowledge grows with experience and that such wisdom cannot so readily be acquired from experts through workshops or in the university classroom. There is also a moral aspect of practical knowledge and that events must be co-constructed by an open and empathetic colleague. Applying this to the APPR teacher evaluation system, the use of rubrics and standardized test scores as a means to evaluate teachers appears to be misplaced. Trying to quantify all aspects of the teaching profession and assign teachers a final evaluation score does not align with this theory nor does it promote wisdom. Instead, this theory supports the notion that evaluators should be experienced professionals with the necessary empathy and context-sensitive judgment to provide meaningful qualitative feedback. Hammerness et al. (2017) describes an alternative to APPR in Finland where evaluations instead consist of a quality assurance model built on two-way dialogue between the evaluator and teacher where the content is co-constructed. This alternative model more closely aligns with Eisner's (1984) theory.

For my research, I will compare teacher evaluation systems in New York State and Finland. For my methodology, I will interview Finnish educators to investigate different types of feedback for teachers and their effects on teacher motivation. The component of framing teaching as an art and craft, and its motivational effects on teaching staff will be especially timely and relevant to

my research. The component of acknowledging the classroom as a domain of practical knowledge where sound judgments about classroom practices are best suited to come from experienced practitioners is also a theory that will inform my work.

Conceptual Framework for My Study

Below is a conceptual framework model that illustrates the various aspects of teacher evaluation systems and factors that may influence teacher motivation to grow and improve. For my study, I crafted interview questions intended to capture the experiences of teachers with these different concepts, and extracted themes that reveal the different aspects of teacher evaluation and their impact on motivation.

Figure 2.1: Conceptual Framework for My Study
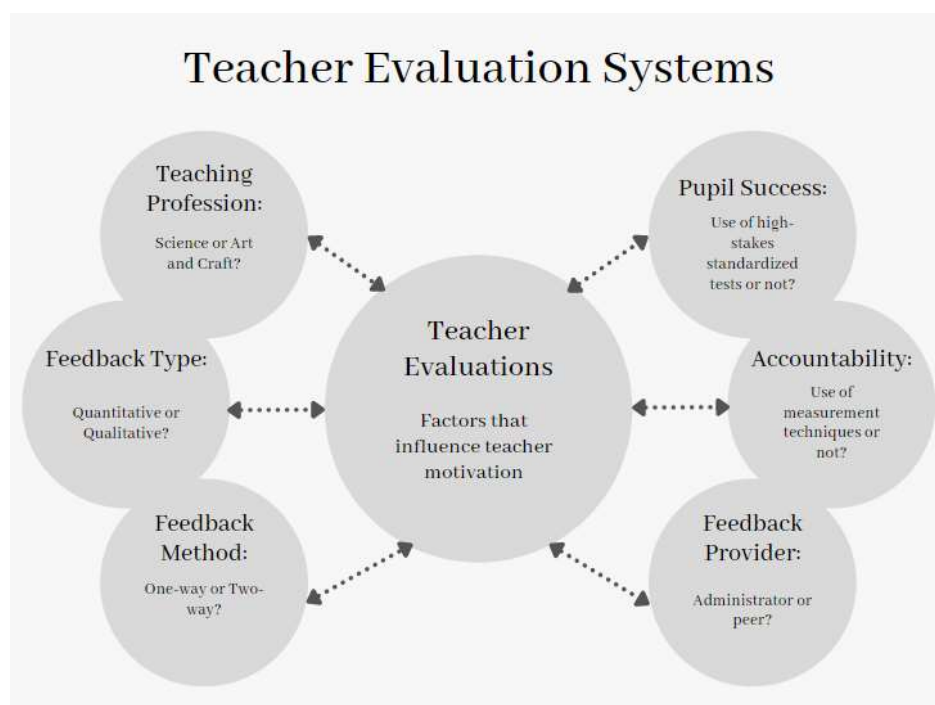
Analysis of Teacher Evaluations in New York State

According to Marzano (2012), teacher evaluations can serve two very different purposes: measuring teachers and developing teachers. Measuring teachers involves determining how

competent a teacher is, whereas developing teachers is about helping them improve. Marzano (2012) goes on to say that an evaluation system that fosters teacher learning will differ from one whose aim is to measure teacher competence. On a scale of 1 to 5, where 1 represents the opinion that measurement is the sole purpose of teacher evaluation and development does not have a purpose, and 5 represents the opinion that development is the sole purpose of teacher evaluation and measurement does not have a purpose, 76% of the 3,000 educators Marzano surveyed selected 4. This indicates that the vast majority of teachers believe that teacher evaluations should be used for both measurement and development, but that development should be the more important purpose.

Danielson's (2011) research was focused on improving teacher evaluations to create better conditions for teacher learning and student learning. She described shortcomings of traditional evaluation systems, such as outdated criteria (usually in the form of checklists), simplistic evaluative comments such as "satisfactory," using the same instruments for both novice and veteran teachers, and also a lack of consistency among evaluators. Danielson (2011) talks about the reasons we evaluate teachers, which include both accountability to the public as well as ensuring teacher quality. Several times in this article, Danielson references the "public's right" to expect high-quality teaching (p. 36). However, one gap in Danielson's research is the impact that of accountability measures has on teachers' motivation to grow and develop. Danielson purports that merging these two purposes of teacher evaluation is the challenge, and to do so calls for the two purposes, accountability and development, to be embedded in the design of the instrument itself. However, is merging the two possible when they are fundamentally two completely different philosophies? Marzano (2012) says the evaluation instruments look very different, depending on the purpose.

Danielson (2011) discusses that in the current evaluation process, the administrator is doing all the work and the teacher is largely passive. This model leads to teachers feeling the process is not valuable nor supportive of their learning. Danielson calls for engaging the learner (teacher) in a process of intellectual engagement. When the teacher and administrator discuss a lesson, the teacher puts the lesson into context for the administrator and together they decide on the teacher's strengths and areas for growth in an effective model. Danielson also says that a consistent definition of good teaching that is understood by all involved in the process is an essential component of models. A thoughtful approach that allows for the necessary time as well as engages teachers in reflection and self-assessment is critical.

Aldeman and Chuong (2014) also made the case that the pre-APPR system gave teachers a score of either satisfactory or unsatisfactory, and referred to this as a 'black and white' way of looking at evaluations (p. 3). The researchers tout the APPR evaluation system, explaining that it is more layered and provides teachers with more feedback. While the increased conversational feedback component promotes greater reflection, negative outcomes are associated with the reformed evaluation system and the assignment of final evaluation scores. Also, it is somewhat misleading that the authors referred to evaluations containing the rating satisfactory/unsatisfactory as black-and-white, because much formative and qualitative feedback was provided to teachers in the pre-APPR system in many cases, alongside a rating of satisfactory or unsatisfactory. This feedback may have been provided conversationally, and was not documented in written form.

According to Smylie (2014), the lack of attention to professional development in teacher evaluation systems is of concern. "Policies governing teacher evaluation systems tend to make only vague and weak provisions for professional development, and they fail to ensure that these opportunities are of high quality and of value in improving practice" (p. 106). Unless teacher

evaluation systems include a professional development component, they will fail to achieve their intended objective of improving teachers' practices.

Challenges in Measuring Teacher Quality

What are the challenges in evaluating teacher quality? According to Labaree (2000), "teaching is an enormously difficult job that looks easy." In fact, the more seamlessly a teacher's lesson appears to flow in the eyes of the observer, the more planning and preparation the teacher has almost certainly completed behind the scenes. In contrast to other professions, Labaree (2000) talks about "control" challenges facing the teacher, such as compulsory education and emotional management of students. Building relationships with students throughout the school year, uncovering their needs and learning styles, and designing lessons to meet those needs is critical work that is often invisible to the observer, and is essential to the visible success of the teacher. Labaree (2000) goes on to say that, "a surgeon can fix the ailment of a patient who sleeps through the operation, and a lawyer can successfully defend a client who remains mute during the trial, but success for a teacher depends heavily on the active cooperation of the student" (p. 228). While instruments exist to measure a teacher's knowledge of their chosen content area, it is much more difficult to measure the intangibles so critical to successful teaching such as instructional delivery, ability to "read the room," and ability to cultivate meaningful relationships with students that motivate them to learn. Darling-Hammond et al.'s (2012) and Meyer's (2016) research supports the claim that teaching is complex work. Gonser's (2020) research illuminated the connection between relationship-building and improving academic outcomes. Students learned best when "teachers saw and heard them as individuals, helped them understand their strengths, and connected what they were learning with their future ambitions" (p. 3).

According to Berliner (2005), it is nearly impossible to test for teacher quality, either in

the certification process or otherwise, particularly under the current time and money constraints facing the field of education. Berliner (2005) makes the case that the assessments currently used to test for teacher quality in state certification processes do not assess the constructs on which they claim to be based. Many times, test questions that contain lengthy introductory materials or are overly verbose correlate very highly with verbal intelligence, which suggests that a different construct of interest is actually being assessed. These test items also have an adverse impact on the performance of nonnative language speakers. In addition, many states also face a labor shortage in the field of teaching, and it has been demonstrated that the passing or cut score on such tests in some of these states appears to be synchronized with the demands of the labor market. This implies that if the tests were made more rigorous, they would likely exacerbate the labor shortage problem. Furthermore, Berliner (2005) suggests that multiple-choice test items often contain more than one plausible choice, leading to confusion on the part of the test taker. "If we genuinely want a highly qualified teacher in every classroom, we should not confuse a highly qualified taker of tests about teaching with a highly qualified classroom teacher" (p. 212).

Another issue with evaluating teachers is that quality means different things to different people. While Harvey and Green's (1993) research took place in the context of higher education, the same underlying principles hold true in K-12 education. One parent may believe that a quality teacher ensures their child learns as much content as possible in a calendar year. Another parent with a child in the same class may believe that a quality teacher makes their child feel loved and welcome at school each day. Simultaneously, that same teacher's principal may believe that a quality teacher is one whose behavior in the classroom results in the least amount of negative parent phone calls to the school. Quality is related to 'processes' or 'outcomes,' and Harvey and Green (1993) looked at widely varying definitions of quality and organized them into five discrete

but interrelated categories: Exceptional, Perfection or Consistency, Fitness for Purpose, Value for Money, Transformation. Without a clearly established definition of quality that reflects the values of all stakeholders, the evaluation process will likely lack credibility.

Popular models of evaluating teachers are also fraught with inaccuracies and inconsistencies (Darling-Hammond et al., 2012). Darling-Hammond et al. (2012)'s research focused on "Value-added models" (VAMs), which are statistical models designed to evaluate student test score gains from one year to the next, and are often promoted as tools to demonstrate evidence of teacher contributions to student learning. Such models are used by school districts in the 20% local assessment piece of the current APPR evaluation system in place in New York State, often in the form of Student Learning Objectives (SLOs). Darling-Hammond et al. (2012) affirmed that gains in student achievement are influenced by much more than any individual teacher, including class sizes, curriculum materials, instructional time, tutors, resources for learning, home and community supports or challenges, individual student needs and abilities, health, attendance, peer culture, prior teachers, prior schooling, differential summer learning loss – which especially affects low-income children, and the specific tests used, which emphasize some kinds of learning and not others and which rarely measure achievement that is well above or below grade level. The research team thus concluded that VAMs of teacher effectiveness are inconsistent and are largely affected by the students assigned to those teachers.

Firestone et al. (2014), found that administrators tended to see the evaluation process as being more accurate and fairer than teachers did. This could be because administrators were observers, while teachers were the ones being evaluated. Concerns were also raised about consistency among raters. Teachers also thought it was important that the evaluators understood the context in their classrooms, and apparently some teachers were given inappropriately-low

ratings because their observers did not understand the reasons for some actions observed. This suggests a concern about teacher trust in their evaluators and more training is probably needed for both teachers and administrators.

In addition, in New York State's APPR evaluation system it is mandatory that student achievement results make up at least 20% of a teacher's final evaluation score, and concerns have been raised about this practice, saying it is unfair and misguided. Foreman and Markson (2015) examined New York State's APPR evaluation system, and its relationship with poverty, attendance rates, per pupil spending, and academic achievement. In this study, it was found that poverty accounted for 60% of the variance in student achievement on high-stakes standardized tests, as opposed to teacher quality. This is problematic for educators because in current APPR practices, teachers are being evaluated and held accountable for factors that are not directly related to their ability as teachers, such as their students' family income.

Furthermore, Foreman and Markson (2015) found the district's APPR system had weak to conflicting correlations with student achievement. The percentage of teachers with a "highly effective" rating was positively correlated with student achievement; however, the correlation was weak. Additionally, the researchers argued that "effective" was the new "ineffective" because so few schools even produced evaluations where teachers were given a final score of "ineffective." It is also of note that as the percentage of "effective" teachers went up, student achievement went down. Overall, the authors made a case that resources would better serve students if they were allocated toward mitigating the negative effects that poverty has on student achievement.

Is APPR a reliable and valid policy in terms of objectively measuring teacher effectiveness? Leonardatos and Zahedi (2014) looked at whether the role of educators has changed, whether there has been any change in autonomy of teachers, whether the culture has changed to

one of mistrust, and whether there was now a diversion of funds from classrooms to corporate profits. The implementation of APPR in districts was analyzed, as well as articles from the press about APPR and also field guidance documents from the State Education Department (SED). The findings demonstrated that the APPR policy is unreliable at times due to its subjectivity and inconsistent implementation across districts and evaluators. In addition, the quality of public education has decreased, a culture of distrust has been fostered, funds have been diverted away from the classroom to external vendors, and 'teaching to the test' has become commonplace in pedagogy.

Additional Challenges in Measuring Teacher Quality

Stronge et al. (2011) conducted a two-phase study that examined the measurable impact that teachers had on student learning and also the instructional practices and behaviors of effective teachers. The results of the study were that top-quartile teachers had fewer classroom disruptions, better classroom management skills, and better relationships with their students than did bottom-quartile teachers. The results of this study are useful in that these are characteristics school leaders and policy makers should be considering when implementing a teacher evaluation system. Discussing classroom management approaches with a teacher is more relevant in everyday practice with students than discussing how to improve students' scores on tests. Stronge et al. (2011) make a case for improving the teaching that occurs daily in the classroom if we are to improve schools, and they argue that to do this we must study what it is that good teachers do to enhance student learning.

Reformers claim that a "science of improvement" is needed to advance the field of teaching (Bryk et al., 2012). Hershberg et al. (2004) affirmed this as well, and said that focusing on growth rather than absolute levels of achievement broadens understanding. Bryk et al. (2012) organized a

group of various stakeholders to gather perspectives surrounding the concept of effective teaching and the ways in which to improve student learning. Different concerns were raised, such as teacher professional development, policy making, high-stakes teacher evaluation, teacher selection, and teacher certification/qualification requirements. The researchers then generated recommendations, such as calling for methods of rewarding effective teaching, using data and metrics to improve how we select, develop, celebrate, and retain teachers, developing practical measurement tools, and detailing what teachers do in practical terms. The researchers claimed that we lack a common language, a shared framework for measuring progress toward higher standards, and that we have weak capacities to learn collectively from effective teachers. These recommendations for improved systems were all cited as a means of closing the achievement gap and also as a means to learn from one another and improving systems through collaboration. Unfortunately, statistical models designed to measure teachers' contributions to student learning, such as Value-Added Models (VAMs), have been shown to have a negative impact in teacher collaboration, which is the opposite of what they were intended to achieve, as the authors were calling for in this article (Darling-Hammond et al., 2012).

While implementing VAMs to determine a teacher's impact on his or her students' achievement is a relatively straightforward process when the course culminates in a state standardized assessment, what about when it does not? Reformers argue there is a growing need for more information about measuring teachers' contributions to student learning, particularly in non-tested subjects and grades (Goe & Holdheide, 2011). In fact, New Jersey Department of Education staff have reported that as few as 20% of teachers will be assessed using student growth data from state tests (Firestone, 2014). Reformers argue that many subjects have this "problem," such as art, music, physical education, world languages, grades K-2, English as a new language,

and special education (Goe & Holdheide, 2011). These are subjects with standards that cannot be adequately measured with a paper and pencil test, and research in this area is very limited. Reformers purport appropriate measures and methods to accurately determine students' growth in every subject and/or grade level are sorely needed (Goe & Holdheide, 2011). Reformers advocate that methods of measuring should also be rigorous, between two points in time, and comparable across classrooms. Some examples of alternative evidence collection methods include curriculum-based pretest and posttest, student portfolios, classroom-based tests, student performance such as in art or music, and other classroom-based evidence that are agreed to mutually by teachers within a given subject.

New York State Teachers' Perceptions of APPR

Hurley (2020) asserts that the use of rubrics is part of a larger culture of evaluation that runs counter to the "inspiration and curiosity needed to create an environment where learning can flourish" (p. 4). The problem with rubrics is how they are used and the ways evaluators interact with them. Those being evaluated become so focused on how to earn a good score or avoid a bad mark that they focus more on how to master the rubric, rather than meeting the intended objectives. This results in those being evaluated missing the larger point of the task. Hurley ultimately calls for abandoning rubrics and re-humanizing education, which is consistent with the theoretical framework outlined by Eisner (1984), and extended by Meyer (2016).

Involving teacher leaders in the evaluation process can influence evaluation meaning and purpose for teachers. Bradley-Levine et al. (2017) conducted a mixed methods study in which teachers were surveyed on their perceptions of the teacher evaluation process. The Teacher Development and Evaluation Model (TDEM) was looked at where teachers were evaluated by school principals, assistant principals, as well as teacher leaders. Teacher performance was

assessed in the areas of student learning and ongoing teacher professional development. It was found that the collaborative professional development, increased structure, and increased transparency were positive aspects that were added by the presence of teacher leaders. However, the presence of teacher leaders was not without problems. Cultivating positive and trusting relationships was still challenging at times. In addition, sometimes teacher leaders were given the cold-shoulder by their peers.

At times, teachers maintain negative perceptions of the measurement aspects of the APPR process. Seymour and Garrison (2016) surveyed physical education teachers on their perceptions of APPR, and found that of the teachers who were polled, 90% disagreed with using measures of student performance to assess teacher competency. Furthermore 80% believed that APPR would not improve the quality of K-12 physical education in New York State.

Assessment is another measurement practice, and it plays a large role in teacher evaluation in New York State. According to Harris and Brown (2009), a tension exists between what "teachers feel is best for students" and "what is deemed necessary for school accountability" (p. 365). The researchers describe four common debates regarding assessment, such as that assessments improve teaching and learning, assessments hold students accountable, assessments can measure the quality of teachers and schools, and that assessments should be rejected completely because they are an invalid means to measure learning. In general, teachers hold complex views of assessment and use different assessments for different purposes.

Harris and Brown (2009) also identified and described seven major purposes for assessment. They were compliance, external reporting, reporting to parents, extrinsically motivating students, facilitating group instruction, teacher use for individualizing learning, and joint teacher and student use for individualizing learning (Harris & Brown, 2009). There are also

four significant pressures that teacher face regarding assessment, including protecting students from harm, fulfilling employer obligations, seeking to improve schools, and helping students learn. It was also emphasized that teachers found it challenging to juggle multiple stakeholders' interests, with the most significant tension being the balancing of school and student needs. The authors conclude that the tensions need to be resolved between teacher and school so that assessment can fulfill its real purpose.

In the case of "expert-driven" Race to the Top initiatives such as reformed teacher evaluation systems and Common Core implementation, Hess & Willey (2018) argue that the biggest mistake of the Obama Administration was the failure to engage their critics. Without an understanding of the real-world barriers to implementation, the initiative was derailed from the outset by parent pushback. While Hess and Willey's (2018) diminishment of expert-driven policy seems a little harsh at times, their concerns about experts ignoring societal complexities are insightful. When people think of an expert, they often think of someone who is all-knowing, and thus should be automatically deferred to in decision-making. But experts can be too close to a situation to remain objective, falling victim to "groupthink" and the inability to listen to criticism. This can leave them vulnerable and unable to see clearly to avert potential disasters (Hess & Willey, 2018). Integrating Harris and Brown's (2009) work as well as Marzano's (2012) research, Race to the Top initiatives such as assessment and teacher evaluations were implemented with a measurement focus, when in practice this was antithetical to the work teachers do each day, which has a growth and development focus. The theory outlined by Eisner (1984) presents a case that science and measurement approaches to education run counter to promoting growth and development, and the researcher argues for an approach of an artist that is rooted in practical wisdom.

Unintended Consequences of APPR

"Not only were our educational leaders devoting their time and energy to matters that are incidental to the real purpose of schools, but our teachers were forced to spend countless hours on meaningless clerical work - hours that should have been devoted to teaching and learning. And, unfortunately, much of this clerical work has survived down to the present time" (Callahan, 1962, p. 178).

Kraft and Gilmour (2016) discuss the purpose of teacher evaluation, and then analyze the role of the principal as the lead evaluator. They discuss the reform movement, and how the role of the principal and the documentation process has changed. They then complete a study of principals and attempt to better understand whether principals feel they are able to promote teacher growth in their newly reformed role as evaluators. Kraft and Gilmour (2016) interview principals and arrive at two important findings. First, on a positive note, they learn that the reformed evaluation rubrics provided a common framework that was very helpful in facilitating conversations around teaching and learning and providing teachers with feedback. However, on a constructive note, giving principals this new role resulted in implementation challenges, such as a decline in the quality of evaluation feedback provided to teachers.

Santoro (2011) discussed what it means to have good teaching and what it means for teachers to experience burnout and demoralization. Santoro (2011) defined both burnout and demoralization and made the distinction between the two terms. Santoro (2011) also discussed how experienced teachers who are fueled by the moral rewards of the profession struggle at times when those moral rewards cannot be accessed.

Through collecting interview data from experienced professionals in the field to try to study the demoralization piece, Santoro (2011) learned that accountability measures such as annual yearly progress, standardized testing, and prescriptive learning standards really got in the way of what teachers would describe as "good teaching." The research collected demonstrated the

destructive influence of high stakes accountability on the quality of teaching and learning. It also revealed that many demoralized and burnt-out practitioners leave the profession, although there are some who are demoralized yet still choose to stay.

Questions were raised regarding whether the new evaluation system would improve physical education. In the past, many physical education teachers evaluated their students using performance-based models, and it was hypothesized that APPR implementation would result in many physical educators now increasing their use of paper and pencil assessments (Seymour & Garrison, 2016). Seymour and Garrison (2016) sampled physical education teachers in New York State to study their attitudes on the state's new teacher evaluation policy. The research questions in the study looked at the type of evaluation mechanisms school districts were using in New York State to evaluate physical educators, as well as whether the APPR system was a sound method for evaluating teachers and their content area.

Seymour and Garrison's (2016) survey results also revealed a tension in the profession. Some teachers chose to focus on fitness tests over pencil and paper tests, even at the risk of an ineffective teacher reading. Others chose to adopt a written test, even though it is less related to the content of physical education but easier to document student growth. Overall, physical education teachers responded with strong reservations about using value-added logic to evaluate physical educators.

It is well-researched that principals are key factors contributing to teacher success, and subsequently student performance (Leithwood et al., 2004; Ni et al., 2018). Because of this, the topic of principal and teacher wellbeing are attracting more attention from policy-makers and supervisors (Sahlberg, 2021). A recent survey of principals conducted by the National Association of Secondary School Principals (NASSP) found that 42% of principals want to leave their position,

and some of the most commonly reasons cited were working conditions, lack of decision-making authority, and high-stakes accountability systems and evaluation practices ("With nearly…", 2020). Leithwood et al. (2004) found that principals were second only to teachers in impacting student learning, and with principal shortages becoming a growing concern in the United States it will soon become critical to remedy the reasons why principals are leaving the profession ("With nearly…", 2020). Of those principals planning to leave their school, heavy workload was expressed as a significant concern. With the increased demands of teacher observations, principals compensate for this by reducing time spent on supporting teachers who need it, providing less feedback, and offloading tasks to others (Curtis, 2012; Firestone et al., 2013; Milanowski & Kimball, 2003). The time required to collect information for teacher observations competes with the time principals need to create working conditions for efficacy (Firestone, 2014). In addition, principals cited their own high stakes metric-driven evaluations as a reason for wanting to leave their positions ("With nearly…", 2020).

Stakeholders have both educational and financial reasons to entice low-performing schools to improve, and because of this school improvement strategies often involve social pressure (Feng et al., 2010). According to Feng et al. (2010), school accountability ratings often correlate with housing values. However, the impact of this accountability pressure on teachers is that they leave the high-pressure school. In fact, in schools that are downward-shocked, which means receiving a lower accountability grade post change then would have happened before, teachers are less likely to be retained than in schools that received no accountability shock. This is especially true for schools that receive a downward-shocked grade of F. Conversely, schools that are not hit with an accountability shock do not experience any significant change to the quality of teachers that leave or stay. This phenomenon requires further analysis in light of teacher burnout brought on by the

worldwide COVID-19 pandemic, where 55% of teaching staff are now saying they will leave teaching sooner than they originally planned, according to a recent poll by the National Education Association (NEA) (Kamenetz, 2022).

According to Feng et al. (2010), school accountability systems cause school leaders to behave differently and school personnel are very responsive to increased accountability pressure. Some unintended consequences of accountability pressure include altering school nutrition programs on testing days to increase the likelihood that students perform well on their exams, suspending students at different points in the testing cycle in an attempt to alter the composition of the testing pool, and reclassifying low-achieving students as learning-disabled so their scores will not count against the school and accountability systems. There is also evidence that teachers are more likely to cheat on exams because of this accountability pressure.

Accountability and Perception of Teachers in New York

There are many countries that outpace the United States in student achievement, yet rarely access the education reform approaches that the United States utilizes (Williams & Engle, 2013). Williams and Engle (2013) studied the ways in which other countries evaluate teachers, namely Finland, Korea, Japan, Singapore, and Ontario in Canada. The researchers identified four primary approaches to accountability across systems: professional accountability, organizational accountability, market accountability, and parental/community accountability. School systems typically favor one mode of accountability, although it is often the case that all are present in some form. The United States system relies heavily on market accountability, whereas "Finland's teacher evaluation system is based almost entirely on professional accountability" (Williams & Engle, 2013, p. 54). In Finland, evaluation processes are consultative and informative in nature. While organizational accountability exists in Finland, its purpose is largely to inform the

professional development of teachers rather than order compliance. Additionally, in Finland, high-stakes standardized testing is not a common practice, and thus does not play a role in teacher evaluation.

Pre-APPR evaluation systems were questioned for their use of binary ratings, such as satisfactory or unsatisfactory, and Arne Duncan is quoted as saying "today in our country, 99% of our teachers are above average" (Glazerman et al., 2011, p. 3). Prior to APPR implementation, reformers argued that a new system of teacher evaluations was needed in order to make the measurement of teacher performance and feedback more rigorous and useful. To accomplish this, reformers argued that a value-added measurement (VAM) was needed in order to determine the impact teachers directly have on their students' growth. The reformers proposed this as a worked solution to a specific administrative challenge of identifying low-performing teachers in need of intensive professional development. They also argued it would set minimal standards for the reliability of teacher evaluations.

Proponents of VAMs, such as Glazerman et al. (2011), Chuong and Aldeman (2014), and others argued that the use of them would result in a more reliable measure of a phenomenon that will yield similar results when replicated. However, since this report was completed in 2011, numerous studies such as Darling-Hammond et al. (2012) have shown that value-added measurements are unreliable and vary greatly from year to year depending on a teacher's students. Additionally, other factors such as poverty correlate more highly with student achievement (Foreman & Markson, 2015). The argument for reforming teacher evaluation systems as part of the Race to the Top initiative was that strengthening evaluation systems would boost student achievement (Ravitch, 2021). However, Ravitch (2021) recently reported that a decade of reforms focused on tougher teacher evaluations has produced no improvement in student test scores.

Prominent classical organizational theory concepts, such as measurement in Taylor's (1967) *Principles of Scientific Management* and systematized technical perfection in Weber's (1947) *Bureaucracy* can be found woven into the current APPR system in New York State, and this is problematic for educators at times. In Taylor's theory, he argues that in the absence of scientific management principles such as measurement, employees will engage in rampant "soldiering," or doing the least amount of work as possible and coercing their colleagues into to doing the same (Taylor, 1967). However, this depiction of employee behavior is not only inaccurate, it is demeaning. Portraying teachers in this way undermines the very premise of the education profession and leaves teachers feeling disrespected, undervalued, and distrusted. In addition, the detailed scoring rubrics districts must use in the APPR process to measure teacher effectiveness, which parallel Weber's theory of Bureaucracy, have a pigeon-holing effect and leave evaluators with limited means in which to capture and score the effectiveness of teachers. The nuanced, intangible, and often invisible aspects of teaching that are so critical to students' success are not adequately captured or measured by such instruments.

Another complication facing school districts is that they are largely funded by local school taxes, which are based on community members' property values. This results in a level of accountability of school districts to the larger community, and many community members struggle to understand the intangible benefits of teaching and at times question what they are "getting for their tax dollars." In *Education and the Cult of Efficiency*, Elwood P. Cubberley is notably quoted in 1916 as saying, "the purpose of the new scientific movement was to create standards so that the efficiency of the work of schools could be determined, demonstrated, and communicated to the public in a language which the community could easily understand" (Callahan, 1962, pp. 97-98). Both communication with the public and reassuring communities that teachers are being held to

high standards are still barriers for school districts today in their pursuit of progress. While the teacher-quality measurement aspects of the APPR process are useful accountability and public communication tools in this regard, those same aspects are also the parts of the APPR process by which teachers are off-put and serve to undermine the growth aspects of the evaluation process. Teachers find it frustrating that they are being "ranked" and "rated" using instruments that only capture narrow, visible aspects of their craft. How can school districts satisfy the accountability demands of the public while also evaluating teachers in a way that does justice to their impact on student learning, as well as motivates them to continue to improve?

Schools have long been criticized for their inability to quantify what their students have learned. In American culture, there is a public expectation of accountability, which can be very challenging in a field where reliable instruments to measure precisely what students have learned and who is responsible for them learning it do not exist. In addition, the vast majority of the public does not possess the requisite knowledge or experience to be considered informed of what teachers actually do. The desire for such instruments has been present in American culture for some time. "In the process of actually attempting to measure efficiency within the schools, educators engaged in a wide variety of activities, but most of the attention was devoted to developing and utilizing 'objective' achievement tests in the language arts and arithmetic and in developing scales for rating the efficiency of teachers." (Callahan, 1962, p. 100). Historically, reports on schools have been questioned for being too theoretical and filled with opinion. As one writer put it, 'Of the text of the superintendent's report one thing should be said emphatically: It must embody facts, not theories, or it is useless.' Why? Because, he said, 'Achievement, not intangible theory, approves itself to the taxpayers and wins their confidence in future measures" (Callahan,1962, p. 156-157).

According to Carrell & West (2010), when evaluations are based on test scores, teachers

increasingly "teach to the test." Although teacher accountability is expected by the public, one problem that arises is that teachers then feel pressured to ensure their test scores fall within the range that is expected of their communities. Carrell and West (2010) conducted a quantitative study that looked at the problems with ways in which professors are evaluated. It was found that evaluations were often based on student perception of how well the course went, and thus professors could inflate student grades or dilute academic rigor in the spirit of improving their own evaluation results from students. It was also found that in introductory courses, while the professors that lacked experience were evaluated highly by students, in future classes whose content depended on that which was taught in the introductory class, students performed worse. However, in introductory classes where professors had much experience and the class may have been more academically rigorous, this resulted in lower student ratings of professors on evaluations, though students actually performed better in subsequent classes even though they evaluated the professor lower. These findings are important because they speak to the behavior of educators being different when they are being evaluated using measurement approaches, rather than when they are evaluated using growth and development approaches.

The media plays a role in portraying teachers to the public. According to Ulmer (2016), the media perpetuates the narrative of a national crisis in teacher quality. This covertly advances the reform agenda by providing a rationale for the restructuring of the education system centered on accountability and the measurement of teacher quality by his or her students' performance on standardized assessments. Newspapers have influenced public opinion of teachers and of the education profession, and the media also serves as a policy actor at times. To combat this, the use of counter-narratives from actual teachers could expand the discourse on teacher quality and offer a more complex and nuanced approach. Furthermore, blogs, Twitter, journalistic columns, and

social media have significantly advanced the voices of teachers and brought a sense of democracy to the media. However, teachers must continue to actively contribute counter-narratives as a means of offering an alternative to the ongoing narrative of low teacher quality. According to Woo (2019), when our society values teaching as a profession and policy systematically supports teachers, we can expect responsible teachers instead of accountable teachers.

An Alternative to APPR: How does Finland assess teacher quality?

Basic Features of the Finnish Education System

The Finnish education system consists of comprehensive schooling, secondary education, higher education, and adult education. Kansanen (2003) states there is also elective pre-school education offered to children aged five and six. Comprehensive schooling consists of nine years for the whole age cohort, beginning at age seven. Secondary schooling consists of either three years of general education in an upper secondary school, or two to six years of vocational education. Classroom teachers handle grades one to six, and they teach all subjects and guide the whole personal development of their students. Subject teachers teach grades seven to nine, and they usually teach one or two subjects. More than half of students pursue upper secondary education for another three years, ages 16 to 18, which culminates in a matriculation examination that is required for students intending to pursue university studies. In addition, fewer than 40 percent of students enroll in vocational education, and some of these programs can serve as pathways into university studies. There are also many students who pursue vocational education following upper secondary school (Ministry of Education, 1999). Finnish students also spend significantly less time in school than students in the United States, and are allotted 15 minutes of recess every hour (Haapala, 2017; Yli et al., 2016).

In Finland, teaching is regarded by its citizens as one of the most prestigious occupations

available, and teacher education programs are highly competitive and available to only the top students through Finnish universities (Kelleher & Kase, 2012). Teacher education is academic and takes place in universities (Kansanen, 2003). Educational psychology and sociology play important roles, and a key feature is that the content of knowledge for pedagogy must be sufficiently broad so that teachers may guide the development of their students as extensively as possible. In addition, the intention of the programs is to closely link theory and practice so that teachers are equipped to resolve everyday teaching problems using the basis of their theoretical knowledge as a framework. To be a teacher in Finland, one has to have a master's degree, be in the top 10 percent of their class, and have two hours of professional development each week. Finnish teachers who are holders of a Master's degree may further their studies and take a PhD examination.

There have been a few problems with Finnish teacher education systems. First, the work of a teacher has always been popular among young women and much less so among young men (Kansanen, 2003). The most common explanation for this is the low teacher salaries. Another explanation for this is that the high academic level of teacher education makes it possible for graduates to apply for other forms of employment in society. In addition, there are a few subjects, such as mathematics and physics, where there is a much lower enrollment in teacher education programs.

Hansén et al. (2012) looks at university teacher preparation programs in Finland, and acknowledges what the U.S. and Finland can learn from one another by comparing policies and practices on a global level. There are many tensions and challenges that exist in teacher education programs (Hansén et al., 2012). For example, to be a teacher in Finland you must have a Master's degree, and that the preparation is largely theoretical and research-oriented in nature. Critics say

that there should be more of a practical component to the preparation model. While some say it might be old-fashioned and in need of reform, the truth is there are flaws with every model, and time has proven that this model largely works. It is hard to argue with the country's success overall, and Finnish education systems remain highly regarded in the world.

There has been a trend in Finnish curricular development towards decentralization (Kansanen, 2003). In Finland, the National Board of Education only sets the framework of the curriculum and leaves room for communities and individual schools to make local decisions with respect to what is taught (Kansanen, 2003; Webb et al., 2004). Curricula is thus school-based with a high degree of local control (Kansanen, 2003; Webb et al., 2004). In fact, teachers are very enthusiastic about planning their own curriculum and a rhetoric of teacher empowerment has become key to the concept of teacher professionalism in Finland (Webb et al., 2004). When governments are overly prescriptive, this undermines public respect for teachers because it conveys a message that "teachers are people who need telling what to do" (Webb et al., 2004, p. 92). Teachers are responsible for planning their own teaching, and may work with other teachers in their school to review each other's work, coach one another, and make curriculum plans across grade levels and content areas (Kansanen, 2003). In addition, Finnish teachers reflect their academic tradition by placing emphasis on a "businesslike" type of behavior in the classroom, as opposed to their American counterparts who tended to be much more dynamic in their classroom behavior (Tirri, 1993).

Three key values in Finnish culture that allow their schools to thrive are trust, collaboration, and wellbeing (Kelly et al., 2018). According to Kelly et al. (2018), "the use of trust-based responsibility over test-based accountability allows for greater teacher autonomy and student-centered assessment and learning experiences" (Kelly et al., 2018, p. 34). This promotes a culture

that fosters support for student learning and social-emotional growth and development. Kelly et al. (2018) also discusses the history of major social reform that began in Finland in the 1960s where high quality and heavily subsidized childcare and early child education programs were instituted nationwide. The impact of these social programs has been far-reaching, and has served to mitigate the adverse effects of inequality that children would otherwise experience early in life, prior to beginning school. As previously mentioned, Finland's childhood poverty rate is only 3.6% versus the United States where our childhood poverty rate is 20.2%. In Finland, education is seen as a way to even out social inequality. In the United States, the focus is on test scores instead of on development of students. Trust, collaboration, and wellbeing are central to the Finnish philosophy, and strong community partnerships where project-based learning takes place that directly supports the larger community are the vessels in which this happens.

Sahlberg (2021) suggests that the world can learn much from Finland's experience with educational change. The conceptualization of teachers and the teaching profession in Finland is very different from that of the United States. In Finland, teachers are regarded with high social esteem and are compensated at a rate that is competitive with other high-paying professions (Sahlberg & Walker, 2021). The process for becoming a teacher in Finland is very rigorous. Finnish teachers are given a great deal of autonomy and their professional judgment is trusted by both administration as well as the larger school community.

There are many reasons why Finnish schools became so successful in the 2000s (Sahlberg, 2021). In Finland, mutual trust between government, society, and teachers is a focus (Snider, 2011). As teacher quality has increased, the Ministry of Education has devolved more authority to schools and teachers are provided autonomy to make educational decisions within a broad framework of goals. Other positive contributing factors to Finland's success include pre-service

and in-service training, working conditions, respect for those entering the profession, low attrition, small class sizes, elimination of tracking, and less time in the classroom compared to other countries (Kelleher & Kase, 2003).

Finland and other high-achieving countries continue to develop great teachers by building in substantial time for regular collaboration among teachers on issues of instruction (Darling-Hammond et al., 2010). As previously mentioned, because Finnish students are in school for far fewer minutes than students in the United States, teachers have more time to plan, collaborate and prepare. Finland is an example of a country that has been successful at decentralizing authority and granting local municipalities, schools, and teachers a high level of autonomy. Aside from a college entrance exam, Finland does not have standardized high-stakes tests. In addition, only a small selected sample of schools participate in the national and international assessments such as the Program for International Student Assessment (PISA) (Woo, 2019). Darling-Hammond et al. (2010) found that responsibility for the evaluation of student outcomes belongs to each teacher and school. Extensive opportunities for ongoing professional learning are supported by leaders with substantial planning and collaboration time in school. This is in contrast to the United States where professional development is more fragmented. In the United States, studies have shown that very few American teachers receive the kind of sustained, continuous professional development that research indicates can change teaching practice and improve student achievement.

Although Finland is regarded as having an excellent education system, Sahlberg (2021) warns that education rankings offer too narrow a view to judge the success of education systems, such as results on the Programme for International Student Assessment (PISA) exam. Sahlberg (2021) also suggests that other countries should be cautious in imitating the Finnish model, due to the unique context and conditions in each individual country. In fact, much of the innovation taking

place in Finland's classrooms surprisingly originates from other countries, notably Germany, Switzerland, and Sweden. Sahlberg (2021) also says more recently England, Scotland, Canada, Australia, and the United States have served as sources of inspiring ideas to enrich teaching and learning in their schools. Finally, Sahlberg (2021) talks about how Finnish schools were very well-performing in the 2000s, however more recently there have been some trends that have educators and policymakers in Finland worried about the future. Notably, scores are dropping, less young people read for pleasure, and equity of education has declined. There is also still much to learn about the disruption caused by the global COVID-19 pandemic.

Finnish schools are becoming increasingly multicultural. Kansanen (2003) states there are increasing numbers of immigrants, in addition to the national minorities of Swedish-speaking people as well as the Sami population. Understanding different cultures is an expectation in Finnish schools, and multicultural aspects of education are emphasized in Finnish curricula.

In addition, the pedagogical issue most frequently raised by Finnish educators is lack of knowledge and skills in supporting students with special education needs in the classroom (Webb et al., 2004). This results in teachers' confidence being undermined and adversely affecting their perceptions of themselves as professionals. There is also an increasingly "anti-school" subculture emerging among students in Finland, according to teachers. This has led to teachers feeling that the work of a teacher is expanding to include new and different roles, such as those of a social worker, father and mother. Furthermore, Finnish children and their parents are being reconceptualized as clients and customers, and are increasingly entitled to participate in the future direction of schools. Thus, an increasingly important feature of Finnish teacher professionalism is to cultivate an ability to relate well to parents and caregivers, to ensure they are well-informed about the work taking places in schools and ways they can be involved in supporting the learning

of their children.

Alternative Practices in Teacher Evaluation: The Case of Finland

Hammerness et al. (2017) discuss the policies and practices that have worked to support an educational context in Finland that is centered upon children. Finland also has a policy context that builds capacity for quality teaching. The assessment of teachers in Finland is focused upon professional development at the individual level rather than on standardized test scores. In Finland, rubrics are not used and there are no documentation processes for evaluating teachers. Instead, feedback is provided face-to-face and is often in the form of informal dialogue with the school leader. This practice is also supported by Hurley's (2000) research on the negative consequences of rubric use in providing feedback.

In an interview, Minister Virkkunen said that teacher evaluation in Finland is based on trust and cooperation with findings used for development (Snider, 2011). Hammerness et al. (2017) assert that a huge part of the success of the teacher evaluation process is that school leaders engage in "management by walking around" in their school buildings and are keenly aware of the work taking place in classrooms between professionals and children. It works because of the closely-networked school community that is created. Some feedback to teachers focuses on key features of teaching, such as personal performance, versatility, initiative, and ability to cooperate. Again, there are no standardized test scores, no value-added data, and no quantitative indicators. In addition, feedback provision takes place in the form of a two-way dialogue where the administrator and the teacher come to agree upon the content.

Other conditions that support the success of Finnish education pertaining to evaluation practices are the labor agreements for teachers that are negotiated collectively, the rigorous university preparation of teachers, and that teachers make a reasonable salary that is commensurate

with other professions (Hammerness et al., 2017). Teachers are looked at with high social prestige and are given significant professional autonomy in schools. In general, teaching is viewed as a service to society and the public good in Finland.

Tirri and Puolimatka (2000) address concerns in teacher preparation in Finland regarding teacher authority in practice and an ability to manage the classroom. In Finnish schools, teachers are educated at universities in a style that is largely academic and research-based, and designed to promote pedagogical thinking through theoretical approaches that can be applied to every day practice by teachers. In this preparation, one challenge is to prepare future teachers to identify and solve real professional dilemmas they will one day face. While those charged with hiring teachers in Finland look at many factors such as a teacher's personality, strong content knowledge, and innate interest in the teaching profession, Finnish teachers actually feel ill-prepared for demonstrating authority in the classroom. This is a potential area of weakness in teacher preparation programs.

Tirri and Puolimatka (2000) describe two types of authority, epistemic (content-knowledge) and deontic (ability to give orders). Finnish education supports the development of epistemic authority in teachers, however deontic authority appears to be lacking. "Teachers need the deontic authority to control the classroom situation and to maintain order" (Tirri & Puolimatka, 2000, p. 159). The authors also touch on a fundamental tension in education where the ideal is for students to be naturally curious about the content and for the need for the deontic authority to be minimal, however the authors also acknowledge that in reality this is rarely the case. It is common for a person to try to exercise authority in a field where their authority does not extend, and also for a person to exercise their authority over people over whom they actually have no authority. Overall, the two key points are that the Finnish teachers are known to be good epistemic authorities

but do not always exercise their deontic authority, and the school should support the development of autonomy. In support of these goals, the authors recommend that to improve teacher education, student teachers should be guided to reflect on every day moral issues in schools. They should also be guided to reflect on the teacher's role and authority in resolving conflict. Eisner (1984) and Labaree (2000) touch on these control issues facing the teacher as well.

Finnish Teacher Perceptions of Evaluation and Assessment Practices

Teacher professionalism in Finland is influenced by notions of teacher empowerment (Webb et al., 2004). In contrast to the United States, Finland is able to leverage social capital from their teachers to achieve objectives. Teachers in countries such as the United States are experiencing burnout and demoralization at a rapid rate, and this is largely due to the high levels of accountability and how such policies stand in the way of teachers accessing the moral rewards inherent in the teaching profession (Feng et al., 2010; Santoro, 2011). Sahlberg & Walker (2021) present a vision in Finland in which teachers are empowered, trusted, and their ideas serve to drive the education system at large. There are seven principles the authors cite as part of the vision. They educate teachers to think, mentor the next generation, exercise professional autonomy, cultivate responsible learners, play as a team, share the leadership, and trust the process. They claim, schools change at the speed of trust, and trust in teachers is guiding principle in the Finnish education system.

Could New York State Learn from Finland?

Factors that Influence Teacher Motivation

According to Firestone et al. (2014), teacher motivation is strongest when individuals feel both competent to carry out their assigned tasks as well as expect that doing so will have the intended effect. Competence alone is not enough, nor is expectancy. In general, individuals who

are internally motivated experience both autonomy and self-efficacy. Extrinsic rewards, such as money, come with challenges such as the limited ability of the supervisor to monitor how the work is accomplished outside of the measurement of the final outcome. Teaching to the test, as well as cheating, are examples of dysfunctional consequences of extrinsic incentives (Firestone et al., 2014; Feng et al., 2010).

Trust between the teacher and his or her evaluator is necessary in order for feedback to be processed and accepted (Arneson, 2014). According to Bryk & Schneider (2003), in order for a school community to function effectively, there must be agreement in role relationships across the school in terms of understanding personal obligations and expectations of others. Any deliberate action that was taken to reduce a sense of vulnerability in others and make them feel safe and secure was shown to build trust in the community. It is important to note that that people typically avoid demeaning situations if they can. Trust also reduces the sense of risk associated with change. Consistency between words and actions of leaders was also an important precursor to establishing trust. Schwab & Zahidi (2020) assert that trust is a central component to the Finnish education system, which has been cited as the most well-developed education systems in the world. In contrast, New York State's APPR evaluation system is based on high accountability that could be described as "command and control," and communicates an underlying lack of trust (Williams & Engle, 2013, p. 54; Woo, 2019). In addition, the term "accountability" is hardly used at all in Finnish literature, particularly in the area of evaluation (Sahlberg, 2021).

In addition to trust, collaboration has become another essential component in navigating reform efforts in schools, particularly at the district level (Daly et al., 2015). Daly et al. (2015) studied the impact of negative relational ties between leaders and how they impact educational systems in efforts to improve organizations. The researchers explored factors that may contribute

to the likelihood that educational leaders form negative professional relationships. In organizations with high levels of social capital, or a strong web of social relationships with high degrees of trust and a culture of innovation, educators were better able to improve outcomes. Additionally, the positive or negative nature of the relationships also correlates with the flow of both information and resources. In organizations with positive relations, information and resources flow freely, versus organizations with negative relations where information and resources either bottleneck or have a gap. According to Firestone et al. (2014), teacher autonomy is one of the critical conditions in schools that maximizes intrinsic rewards, thus fostering teacher motivation.

Organizations have been examining the pros and cons of merit-based pay models for many years (Kelleher & Kase, 2012). While this may serve as one way to extrinsically motivate teachers, there are several potential negative consequences of such models. Teachers echo concerns that if names of teachers who receive merit-based pay become public, relationships will break down and feelings of resentment and embarrassment will result. Other concerns with merit-based pay include increased parental legal action for lack of access to high-performing teachers, increased demand for teaching methods that result in higher test scores, evaluator bias and favoritism, and flight from teaching in low performing school districts. Kelleher and Kase (2012) assert there is no state or national agreement on measuring merit in teaching and the reliability and validity of judgments about performance when it comes to student achievement data. As previously mentioned, Finland does not utilize incentives-based pay in its teacher evaluation system, and instead relies heavily on self and peer appraisal.

Mintz and Kelly (2021) looked at aspects of teacher motivation as they pertain to the APPR teacher evaluation policy in New York State (NYS). Mintz and Kelly (2021) interviewed five high school Living Environment teachers whose courses culminated in a NYS Regents exam, as well

as the five supervisors who were responsible for evaluating these teachers. An important question was how to balance the need for intrinsic motivation with the external accountability expectations of the school community. Future recommendations for teacher evaluation reform include educator involvement when creating new guidelines, and use of test scores being limited to a means of focusing on problematic content. Additionally, increased localized control, peer learning opportunities, and the adoption of more reliable measures were also recommended to address the unfairness in the APPR process. On a good note, teachers appreciated the dialogue and positive feedback from their supervisors as by-products of the APPR process. This aligned with the findings of other researchers in that the dialogue and common language aspects of APPR support teacher growth and development (Danielson, 2011). Dialogue is also a critical feature of the teacher quality assurance model in Finland (Hammerness et al., 2017).

As Marzano (2012) said, teacher evaluations can either serve to measure teachers or develop teachers, and these evaluation systems look very different. The APPR system is measurement-driven and the Finnish quality assurance model is development-driven. Communities of practice is a concept that is gaining momentum in the field that has been shown to be intrinsically motivating for teachers to improve (Wenger & Snyder, 2000). Communities of practice are groups of people informally bound together by shared expertise and passion for a joint enterprise. An example of a community of practice in the field of education would be a team of Algebra teachers who enjoy working together and collaborating on curriculum and instruction. Communities of practice are motivating, naturally sustainable, and have been shown to improve organizational performance.

Wenger and Snyder (2000) tout the "organic, spontaneous, and informal nature" of communities of practice, and also how they are inherently sustainable due to the intellectual bond

that community members share. The participants in these communities of practice learn together by focusing on problems that directly relate to their work. This includes seeking either in-person or electronic opportunities to interact, depending on the time or location constraints of community members. According to Gonser (2020), finding time and head space for reflection is challenging, but absolutely essential to good teaching. The reduced instructional minutes and embedded professional development in the Finnish education model is an approach that is supported by these findings (Darling-Hammond et al., 2010; Yli et al., 2016).

Professional development models such as communities of practice are very appealing to leaders because they have been shown to lead to professional growth, which leads to the betterment of the students they teach and the communities they serve (Wenger & Snyder, 2000). When school administration supports the cultivation of communities of practice, teachers feel valued and respected, thus supporting the sustainability of the communities of practice and positively impacting the school culture at large. Teachers' professional choices are being honored without question, thus making room for creativity and innovation. Presently in the APPR rubric model, teachers are guided to focus on how to get a "highly effective" score, that the majority of their energy goes into satisfying the measurement aspects of evaluation (Marzano 2012; Feng et al., 2010).

Wenger and Snyder (2000) discuss using nontraditional methods to measure value, and they argue that assessing the value of a community of practice is best accomplished by listening to members' stories in a systematic way. This is consistent with Finnish practices of "management by walking around" and two-way dialogue between the evaluator and evaluatee where they come to agree on the content of the feedback (Hammerness et al., 2017). Wenger and Snyder (2000) argue that the way to remedy the conundrum of appraising in communities of practice is to

systematically gather anecdotal evidence.

Another reason why the communities of practice model serve to promote growth and development is because it supports the important notion of practical wisdom, a concept first described by Aristotle that is sorely needed in all organizations, including schools (Schwartz, 2011). Practical wisdom, or *phronesis*, is the ability to perceive a situation, have the appropriate feelings or desires about it, deliberate about what is appropriate given the circumstances, and finally to act accordingly. Aristotle argued that ethics was not mainly about establishing moral rules and following them, it was also about performing a particular social practice. Practical wisdom is not something that can be cultivated through supervisory micromanagement. While a musician can teach a student the mechanics of playing the saxophone, the skill needed for the student to perform improvisational jazz is largely analogous to the kind of skill needed to develop of practical wisdom, only practical wisdom is not an artistic skill but a moral one.

To develop practical wisdom in teachers, they must be given the freedom and opportunity to make a moral choice. When supervisors provide teachers with the autonomy and opportunities to make these choices, such as in communities of practice, teachers not only feel valued and appreciated, but practical wisdom is nourished (Schwartz, 2011). "The rules and incentives that modern institutions rely on in pursuit of efficiency, accountability, profit, and good performance can't substitute for practical wisdom. Nor will they encourage it or nurture it. In fact, they often undermine it" (Schwartz, 2011, p. 5). Practical wisdom is an interactive process that requires judgment and is cultivated during peer-to-peer learning, such as that which takes place in communities of practice.

Bridging the communication gap between what works in improving schools and what the community's accountability expectations are of schools is no easy task in New York State. Wenger

and Snyder (2000) say that because the primary output of communities of practice is knowledge and is intangible, this might sound like another 'soft' management fad. This aligns with the problem Callahan (1962) identified as to why efficiency models remain so popular: Communities expect accountability of schools. The thought of a lack of supervision and demonstrable accountability measures is uncomfortable and unsettling to community members in New York State, whose taxes largely fund school budgets (Feng et al., 2010).

A common assumption in education is that imposing sanctions will motivate educators to perform at higher levels and focus attention on student outcomes (Finnigan & Gross, 2007). However, leaders should exercise caution because low morale has the potential to undermine teachers' responses to these measures. Finnigan and Gross' (2007) study also reaffirmed previous research that has been done that argues the resultant stress and human costs of accountability policies can have counterproductive results, particularly in schools that persistently struggle. This information suggests that policy makers take heed when developing accountability policies that excessively rely on threat or pressure to motivate staff in low performing schools without also providing significant targeted support. In addition, the implementation of support must be more focused on instructional content to minimize the effect on teacher morale.

There is a strong theoretical framework on motivation, and the two concepts that are related to it are the *relative value* an individual places on outcomes and also a person's belief about the likelihood that his or her efforts will result in a desired outcome, or *expectancy* (Finnigan & Gross, 2007). Incentive theory is also discussed, including expectancy, underlying motives, and the interpretation of incentives influence teacher motivation. In addition, a critical interplay between morale and motivation exists and motivation decreases for teachers who work in schools that struggle the most. Teachers are keenly aware of the factors that contribute to student success that

lie outside of the control of the teacher, such as poverty, and being judged unfairly in this way leads to burnout (Santoro, 2011).

Gigante and Firestone (2007) studied teacher leader impact, and found four key resources whose availability will increase the likelihood that teacher leaders will be able to engage in developmental tasks. The resources were adequate professional development time, verbal administrative support, highly trusting relations with the teachers to develop, and teacher leaders' own experience coordinating professional development activities. Developmental functions were found to facilitate teacher learning, whereas support functions had no effect - though they did help teachers do their work. When given access to these resources, time in particular, teacher leaders engaging in developmental functions facilitate education change and improve teaching (Gigante & Firestone, 2007, p. 323).

Potential Limitations to Cross-Cultural Learning

While much can be gleaned from the Finnish approach to teacher evaluation and its practices that foster intrinsic motivation in teachers, Sahlberg (2021) suggests that there are reasons why one should exercise caution in copying the Finnish case elsewhere, such as in New York State. As previously mentioned in the introductory chapter, Finland and New York State are distinct in many ways. Some of these key differences include their governance, geography, demographics, and culture. Thus, even if policymakers New York State were to reform APPR to "match" the Finnish approach to teacher evaluation, there are several factors at work, both structurally and culturally, that could potentially undermine its effectiveness in New York State and elsewhere (Sahlberg, 2021).

Summary

A review of the literature has shed light on the flaws in metric-driven teacher evaluation

systems, particularly those where teachers are evaluated on the basis of their students' test scores. They are demotivating to teachers and ultimately undermine the goal of continuous improvement within organizations. APPR has now been established in New York State for over 10 years. We have learned that it is difficult to test for teacher quality using standardized metrics (Berliner, 2005; Darling-Hammond et al., 2012). Other countries, such as Finland, have found ways to appraise teacher quality that are peer-to-peer and support ongoing teacher growth and development. Leaders can facilitate peer learning and nurture communities of practice by providing support and resources, such as professional development time (Wenger & Snyder, 2000).

In this chapter, I provided a review of the relevant literature surrounding APPR in NYS and teacher evaluation systems in Finland. An important gap in the existing literature is that information can be gleaned from a side-by-side comparison where APPR is compared with an alternative system, such as Finland, and my research will contribute to this by providing further insight and highlighting the differences between teacher evaluation systems. An overview of the specific methodology for my study will be provided in chapter three.

CHAPTER 3: METHODOLOGY

Introduction

This qualitative study set out to understand potential alternatives to APPR, such as the case of Finland. Additionally, this study set out to understand different types of supervisor feedback and their effect on teacher motivation. In this study, I interviewed educators in New York State and Finland on their lived experiences being evaluated as teachers. In addition, I interviewed supervisors in New York State and Finland on their experiences as evaluators of teachers. This analysis was important because while much was known on Finnish school systems and how their guiding principles include trust, collaboration, and wellbeing, little was known on the types of feedback teachers were provided by their supervisors on their teaching and how this contributed to their personal motivation to grow and improve. In addition, a side-by-side comparison of New York State's APPR teacher evaluation system and the Finnish teacher evaluation system offered potentially constructive insight for U.S. education policy. Finland was chosen for a comparison with New York State because of its international success on the PISA exam, and because Finland is regarded as having an excellent education system (Sahlberg, 2011). Finland also made for an insightful comparison due to its education policies and practices being rooted in a professional accountability model, as opposed to being rooted in a market accountability model such as that which exists in New York State (Williams & Engel, 2013).

This chapter explains the methodology of the study in detail. It contains the research questions, study context, research design, description of the sampling methods, data collection and coding procedures, key constructs, ethical considerations, statement of positionality, and the data analysis process. A summary is also provided at the end of the chapter.

Research Questions

As previously mentioned, this qualitative study set out to understand potential alternatives to APPR, such as the case of Finland. Additionally, this study set out to understand different types of supervisor feedback and their effect on teacher motivation.

Research Question #1: How do teacher evaluation systems in New York State compare with those in Finland at both the structural and cultural level?

Research Question #2: How do teachers in New York State and Finland view the contribution of teacher evaluations to their motivation and professional growth?

Research Question #3: What are the policy implications resulting from the above comparison?

Supplemental Research Question #1: How satisfied are teachers and supervisors with teacher evaluation practices in Finland versus New York State?

Supplemental Research Question #2: To what degree do teacher evaluation practices contribute to professional growth in Finland versus New York State?

Study Context

In this study, I conducted in-depth, semi-structured interviews of teachers and supervisors in New York State and Finland on the topic of teacher evaluations. The interview questions I developed were informed by both the literature, as well as a field visit I made to Helsinki and Espoo in Finland. I asked the teachers and supervisors questions that directly addressed my research questions, which are outlined above. The in-depth format was selected so that participant responses included detailed information about the phenomena being studied, i.e., teacher

evaluations. I selected a semi-structured interview format to provide both the structure needed to guide the interview process and address the research questions, as well as the necessary flexibility to adapt the conversation to changing circumstances if needed. These interviews took place via Zoom, and this was due to logistical concerns of interviewing human subjects in locations that are quite a distance from me as the researcher. At the time of the interviews, I was in Albany, New York, and the teachers and supervisors I interviewed were in regions across New York State and Finland. Each interview lasted about an hour in duration.

Research Design

In this qualitative study, my planned research design was to conduct semi-structured interviews of 8 teachers and 3 supervisors from New York State and Finland. I used a set of interview questions that provided opportunities for teachers to share their personal experiences in being evaluated by their supervisors. I also used an additional set of interview questions that provided opportunities for supervisors to share their personal experiences in evaluating teachers. For this study, I used a qualitative method due its strength in understanding teachers' and supervisors' personal experiences in receiving and giving feedback. An in-depth, semi-structured interview process was also the best possible research approach to afford me the opportunity to gain key insight into which specific types of supervisor feedback motivated teachers to grow and develop in their profession. A qualitative research inquiry best provided information on how satisfied teachers and supervisors were with evaluation processes, and how much evaluation practices contributed to the professional growth of teachers.

I interviewed New York State and Finnish teachers and supervisors who have experienced teacher evaluations. This study set out to learn more about "what" teachers and supervisors have

experienced and "how" they experienced it in New York State and Finland. In my literature review, I presented a solid foundation of "what" the teachers and supervisors have experienced, however little has been studied that provided specific insight on "how" they experienced it.

Creswell and Poth (2018) cite five potential interpretive frameworks researchers can utilize and outline the associated philosophical beliefs for each. In this qualitative research, I employed a social constructivist interpretive framework. This means that through the lived experiences of teachers and supervisors, multiple realities were constructed. In addition, reality was co-constructed between the researcher and researched, and it was shaped by individual experiences. Individual values were honored and negotiated between individuals. A statement of positionality of the researcher is provided in this chapter to make clear the identity, life experiences, and values of the researcher a part of understanding this co-construction and interpretation.

Target Population and Sampling

Much is known about APPR in New York State and teacher perceptions in terms of clarity, practicality and motivation (Mintz & Kelly, 2021). However, in order to better understand the range of possibilities for teacher evaluation systems, and to complete a side-by-side comparison, interviews of New York State and Finnish teachers and supervisors were conducted on the topic of teacher evaluations and supervisor feedback.

My sampling procedures were that teachers and supervisors to be interviewed were purposefully chosen based on their location (Finland or New York State), the level that they teach (secondary) and the language they speak (English). I chose to study teachers and supervisors in Finland and New York State because that was the basis for the comparison in my study as outlined in my research questions. I narrowed my sample of teachers and supervisors to the secondary level

to focus the sample.

It was also important that the teachers and supervisors I interviewed were English-speaking because I was an English-speaking researcher, and conducting interviews in Finnish or any other language I did not fluently speak would present challenges in the communication and interpretation of the data. Meaning could have been potentially lost in translation. In addition, a language barrier between the interviewer and the interviewees could have made for an awkward interview experience. Instead, selecting the target population to be New York State and Finnish educators who are English-speaking made for a comfortable environment during the interview process where I could clearly understand what participants were communicating and was able to smoothly adapt to potential changing directions or circumstances in the conversation during the interview process.

Earlier in 2022, utilizing professional contacts of my dissertation committee members, I was able to communicate with Finnish educators to help set up an initial field visit to identify interviewees for this study. After a few successful correspondences through email, phone, and Zoom video conferencing, I was able to successfully book a trip to Finland in May 2022 where I visited three schools in Helsinki and Espoo.

To recruit participants, I had separate procedures for educators in New York State versus Finland. For the Finnish teachers and supervisors, I expanded upon the connections that were fostered during my trip to Helsinki and Espoo and reached out to these contacts. I met some of my potential interviewees during the trip and established a minimal level of rapport. This helped ease the awkwardness that can be present at times when a researcher interviews human subjects where no prior trust had been established (Patton, 1990).

For the New York State teachers and supervisors, I reached out to professional contacts of

mine in the Capital Region in New York State who could disemminate my recruitment email to their respective email distribution lists. All participants were provided with a consent form that provided adequate information about my study. With their consent, participants were interviewed on Zoom. In total, I formally interviewed 8 New York State and Finnish teachers and 3 New York State and Finnish supervisors in this study. Transcripts of the interviews were sent to participants shortly thereafter. No further contact was made with participants.

For the Finnish educators, while it would have been ideal to randomly select a more representative sample of Finnish educators and their supervisors at the secondary level across the entire country, the limitations of the language barrier between the researcher and the human subjects, as well as the potential for awkwardness in the interview justified a more purposeful sampling process as the best available option. There was value in the prior connections made between the researcher and the human subjects during the trip to Helsinki and Espoo. Without a previously established rapport between the researcher and the human subjects, it was possible that I would have struggled to recruit the necessary participants to interview and complete the study.

Although there is already a lot of knowledge on the APPR teacher evaluation system in New York State, fresh data was collected from educators in New York State in this study. This is both to do a true side-by-side comparison, as well as to better support the possibility of these results being generalizable in the future. Collecting fresh data also allowed for the opportunity for educators to share their experiences with teacher evaluation in recent times, and of note, following the COVID-19 pandemic.

Data Sources, Collection, and Organization

This study of teacher evaluations used a qualitative approach, where either evaluating

teachers or being evaluated as a supervisor were the phenomena being studied. My planned data sources were data collected from the individuals who have experienced the phenomenon (Creswell & Poth, 2018). These data were collected using in-depth, semi-structured interviews of 8 teachers and 3 supervisors from New York State and Finland. Interviewing multiple participants provided an understanding of the common experiences of the participants. In addition, the study instruments consisted of one set of interview questions for teachers, and another for supervisors. Due to the cultural differences between human subjects in New York State and Finland, the study instruments were reviewed by a researcher at the University at Albany who speaks both Finnish and English, and who travels to Finland regularly. This individual was able to review the study instruments to ensure they were worded such that they allowed for smooth communication between the interviewer and interviewees during the interview process.

My data collection procedures consisted of the interviews being conducted and transcribed through the Zoom platform. I also utilized the Otter.ai extension for the Google Chrome browser to transcribe the interviews. The interviews were semi-structured and the style was casual. The transcriptions were analyzed using a qualitative data analysis methodology where I went through the data and highlighted "significant statements," or statements that provided insight of how participants experienced the phenomena (Creswell & Poth, 2018).

Key Constructs and How they were Gauged

The construct I gauged was the shared meaning of the phenomena between the participants in the study (Creswell & Poth, 2018). One unit of analysis was teachers, and the other was supervisors. In order to simplify this study, the sample consisted of New York State and Finnish teachers and supervisors at the secondary level. The goal of qualitative research is for the

researcher to construct findings based on the data collected (Terrell, 2016). The constructs to be gauged were the New York State and Finnish experiences of either being evaluated as a teacher, or of evaluating teachers as a supervisor. These constructs were measured by conducting interviews, analyzing the data, and extracting themes. A final step in data analysis was to make an interpretation of the data and assign meaning to the findings in order to clarify what lessons were learned from the inquiry (Lincoln & Guba, 1985).

In-depth, semi-structured interviews were the best means in which to collect data on the lived experiences of human subjects, and subsequently the best means in which to answer my research questions (Creswell & Poth, 2018). As Brinkmann and Kvale (2018) indicated, "If you want to know how people understand their world and their lives, why not talk with them?" (p. 2). Using the information participants provided regarding these shared phenomena best got to the heart of what this study was intending to explore, namely, what were the New York State and Finnish experiences like, and also what types of feedback were motivating to teachers to try to improve their craft.

Validity and Reliability

Qualitative validity means that the researcher checks for the accuracy of the findings by employing certain procedures (Gibbs, 2007). According to Creswell (2013), there are eight primary strategies for establishing qualitative validity, and it is recommended that researchers employ at least two of these strategies to check the accuracy of findings. The eight strategies are triangulation, member checking, rich/thick description, reflexivity (i.e., clarifying researcher bias), negative case analysis, prolonged engagement, peer review, and external auditing.

To validate the accuracy of the information in this qualitative study, I utilized five of the

eight strategies: rich/thick description, prolonged engagement, member checking, reflexivity, and negative case analysis. First, I provided a rich and thick description of the study context, and the research findings. This was to help the reader very clearly visualize the setting and details of the information being shared by the interviewees in New York State and Finland regarding the phenomena. Second, in addition to providing rich/thick descriptions I also shared information about the trip I took to Helsinki and Espoo, and how I participated in prolonged engagement with the Finnish teachers and supervisors in their home country, Finland, while on that trip. Third, following the interviews, I read, formatted, and edited the interview transcriptions carefully. For any unclear portions, I conducted "member checking," where I confirmed the meaning and information with the teachers and/or supervisors. Fourth, I demonstrated reflexivity by acknowledging my own biases as a researcher in a statement of positionality. Finally, when I discussed my findings, I considered negative case analysis of alternative viewpoints as part of the discussion.

An important consideration in qualitative research is for the researcher to maintain objectivity when collecting and interpreting interview data. The researcher must keep personal experiences in check. Regarding reflexivity in qualitative research specifically, Creswell and Poth (2018) highlight the importance of considering the personal experiences of the researcher with the phenomenon that is being studied. An essential step is that the researcher fully describes his or her experience of the phenomenon, and then set aside, or bracket, those personal experiences so that the focus can become the participants in the study. While the personal experiences of the researcher cannot ever be set aside completely, bracketing is nonetheless the best possible practice to promote validity in the study.

Qualitative reliability indicates that the researcher's approach is consistent across different

researchers and different projects (Gibbs, 2007). Given the steps taken to plan for the study and address the potential limitations, it is anticipated that if this study were replicated by another researcher with a different sample of teachers and supervisors, the results would be similar and the findings generalizable.

Positionality Statement

A core feature of conducting qualitative research is acknowledging the position of the researcher. The researcher in this study practiced reflexivity, which was to actively engage in critical self-reflection about potential bias (Creswell, 2013). This was necessary since the researcher was the primary "instrument" for data analysis (Watt, 2007). Including my statement of positionality helps the reader understand the way I look at things and make assumptions as the researcher, and admits that no research can ever be 100% objective. I acknowledge that the same information could have different meaning for someone of a different identity. My background, experiences, and identity in society as it relates to this dissertation is shared below in an effort to clarify my lens and assumptions as a researcher, and provide insight as to how realities were co-constructed between the researcher and interviewees.

The following is a statement of my positionality: I, Lindsay Tresansky, am a building principal at a suburban public middle school in the Albany, New York area. The student body is predominantly Caucasian; however, each year it becomes increasingly both culturally and socioeconomically diverse. I previously served as a high school associate principal and mathematics supervisor in the same school district for five years. I was also formerly a high school mathematics teacher in a neighboring public school district for eight years. I have over 15 years of experience in the field of education. I am also a mother of two children, ages 7 and 9, who attend public schools in the Albany, New York area. I have lived in the Albany, New York area my whole

life and grew up attending public schools.

My background and experience were relevant to this dissertation because of my own work experience with the teacher evaluation process, both as a teacher who was evaluated for many years and as a supervisor who has completed hundreds of evaluations of teaching staff. I served as a teacher from 2008-2016, and it is important to highlight that from 2008-2013 I was evaluated using a local teacher evaluation process and from 2013-2016 I was evaluated using the APPR teacher evaluation system in New York State, which was implemented in all public schools across the state beginning in 2013 as part of the federal Race to the Top initiative. I transitioned to an administrative role in 2016 and have seven years of experience serving as an evaluator of teaching staff across many content areas and grade levels, and have done so using the APPR teacher evaluation system in New York State. My research was informed by a commitment to recognizing the roles that evaluations and supervisor feedback play in the motivation of teaching staff to grow and develop professionally, as well as a commitment to the continuous improvement of schools.

Ethical Considerations

Maintaining the highest possible standards of ethics and integrity was a priority in this study. The rights of all human subjects were upheld. As the researcher, I completed Institutional Review Board (IRB) training on Human Subjects (Investigators, Advisors), as well as the required, elective, and supplemental modules through the Collaborative Institutional Training Initiative (CITI) Program on January 28, 2022. I also obtained approval from the Institutional Review Board (IRB) of the University at Albany before I proceeded with the data collection phase of my study, including the submission of the supplemental paperwork for conducting international research.

Participation of this study was completely voluntary, and verbal consent was required from

all interview participants before commencing with data collection. Prior to obtaining consent, participants were informed of the purpose of the study and their role in its completion. As data was collected, the identities of the participants, as well as any personal identifiable information was kept confidential and stored in a secure location. At the completion of the study, all notes and coding information linking participants with this study was destroyed. When presenting the results of this study, pseudonyms of the participants were used to maintain confidentiality, such as using nicknames for the participants.

Trip to Finland

I had the incredible opportunity to take a trip to Finland in May, 2022. I was able to plan this trip through collaborating with a professor at the University at Albany who spoke both English and Finnish. I visited three schools in Helsinki and Espoo, where I was able to tour the school buildings, meet teachers and principals, observe classes, and interact with students.

Visiting Finland in person afforded me the opportunity to experience the culture in ways that video conferencing and reading books could not. For example, prior to my visit I read that many Finnish students were bilingual and could speak both Finnish and English. However, when I visited a Finnish classroom and introduced myself to the students, a third-grade male student proudly shared with me, "I speak Finnish, English and a little Spanish." It was really great to see his confidence, and it told a story about Finnish education in ways that may not be so readily captured in books.

When conversing informally with teachers at one of the schools, I recall being asked what I was studying at the university. When I answered, "teacher evaluations," the teachers had somewhat confused expressions. I then explained the general process we use in New York State

to evaluate teachers, and they were in disbelief. This is because there truly is no such system in place in Finland where an administrator goes into a teacher's classroom to observe and provide the teacher with feedback. The concept of teacher evaluation was very foreign to the Finnish teachers, and this speaks to the high degree of autonomy and trust that teachers are afforded in Finland.

By completing this initial field visit, I was able to communicate with Finnish educators to identify potential interviewees for this study. Had I not been introduced to a few key connections in Finland, I would likely have not received any responses from teachers or supervisors to my request for interviews.

Limitations

There are many inherent limitations in social science research, and in this study, there were a few limitations to acknowledge that could potentially stand in the way of the results being generalizable. First, according to Creswell and Poth (2018), the preferred number of individuals to be interviewed in a qualitative study would be 5-25. In this study, I only interviewed 8 teachers and 3 supervisors, and this was due to the limitations in identifying Finnish teachers and supervisors who speak English, and who also consented to being interviewed by a researcher in another country. Conducting more than this would be ideal, and I certainly strived to interview as many as I could, however realistically the best available option landed at a sample size of 8 teachers and 3 supervisors. In addition, a simple random sample would likely result in a more representative sample across gender, race, etc., however this study was limited to purposeful sampling for logistical reasons.

In addition to the potentially small sample size and limited sampling method, the

interviews were conducted via Zoom video conferencing. This was due to the logistical challenges of interviewing human subjects in distant locations from the researcher, such as regions across New York State and Finland. It is hard to say what effect, if any, the Zoom video conferencing format had on the participants in the study. Sometimes interactions on Zoom are very comfortable because both the researcher and the human subject can plan to participate in the interview in an environment that is comfortable, such as their own homes or offices. However, a potential downside to using Zoom was that the two-dimensional nature of video conferencing may have led to a less formal environment where the participants behave and interact differently than they would otherwise if the interviews had taken place in person. Another potential downside of the two-dimensional nature of video conferencing was that both the researcher and the human subject were somewhat deprived of the ability to read body language and/or pick up on subtle, physical cues that might otherwise be present in a three-dimensional in-person interview experience. Being deprived of this potential opportunity to sense or perceive information may have impacted the interview process in that researcher or interviewee might have responded differently to a particular stimulus had it been present in a three-dimensional space.

Data Analysis

In-depth, semi-structured interviews of teachers and supervisors were conducted via Zoom video conferencing. The transcription feature on Zoom, along with the Otter.ai Google Chrome extension were employed to transcribe the interview dialogue into written data. The data was then be edited for both typographical errors as well as formatted, as needed, to accurately read like the dialogue that took place. For example, each speaker was identified next to the text that was spoken during the interview, and the transcription was edited to reflect a dialogue format, similar to that of a script for a play. This was to provide the reader with an authentic sense of how the interview

took place. This was necessary since the researcher was the primary "instrument" for data analysis (Watt, 2007). In addition, the qualitative research validity strategy of "member checking" was employed to clarify meaning with the participants regarding the phenomena being studied (Creswell, 2013).

The data were coded and stored in a secure location via Microsoft Office. I chose Microsoft Office (i.e., Microsoft Word and Excel) because of its strength as a software to best organize my data and facilitate the data analysis process. In Microsoft Excel, I entered the transcript text and used different spreadsheets to sort and structure the data. I was also able to use other organizational tools to assist in coding the data, such as formulas and highlighting. Once the data were coded and sorted by content, they were grouped into categories and subcategories. "Significant statements" then emerged, and were grouped into "clusters of meaning," or themes (Creswell & Poth, 2018, p. 79). These coding and storage procedures were chosen for their strength in organizing the data and facilitating the data analysis process.

The transcriptions were then analyzed using a qualitative data analysis methodology where I went through the data and highlighted "significant statements," or statements that provided insight of how participants experienced the phenomena (Creswell & Poth, 2018). This process was employed to answer both the three primary research questions, as well as the two supplemental research questions. It was employed for both the data collected from interviewing teachers, as well as the data collected from interviewing supervisors. In addition, the two data sources were brought together and analyzed for additional "significant statements" to form "clusters of meaning" in both data sets. A final step in data analysis was to make an interpretation of the data and assign meaning to the findings in order to clarify what lessons were learned from the inquiry (Lincoln & Guba, 1985). As part of this, differences between the New York State and Finnish systems were analyzed

at both the structural as well as the non-structural levels. Attention was paid to differences at the cultural level by looking at the unstated, shared assumptions, and it was insightful to find out what was implied, or taken for granted at times in education in looking at educator perceptions in both New York State and Finland.

Summary

This qualitative study examined potential alternatives to APPR, such as the case of Finland. Additionally, this study sought to understand different types of supervisor feedback and their effect on teacher motivation. I used an interview protocol of 13 questions each to conduct interviews via Zoom of 8 teachers and 3 supervisors from New York State and Finland, who speak English. I organized the data, and then grouped the data into categories and subcategories. "Significant statements" emerged and were then grouped into "clusters of meaning," or themes (Creswell & Poth, 2018, p. 79). Attention was given to plan to extract any unstated, shared assumptions at the cultural level. While there were some limitations in the size and sampling procedures in this study, it was informative to analyze the data from the interviews of the teachers and supervisors. Completing this study provided the desired key insight into the established research questions and made the Finnish case of teacher evaluation clear so that a comparison with New York State could be made. The next chapter contains the results of the study. For organization purposes, chapter four will be organized by region. It will also be organized by the five emergent themes.

CHAPTER 4: FINDINGS AND RESULTS

Introduction

In this study, I interviewed educators in New York State and Finland on their lived experiences in being evaluated as teachers. In addition, I interviewed supervisors in New York State and Finland on their experiences as evaluators of teachers. This research is important because, while much is known on Finnish school systems and culture, little is known on the types of feedback teachers are provided by their supervisors on their teaching and how this contributes to their personal motivation to grow and improve. In addition, a side-by-side comparison of New York State's APPR teacher evaluation system and the Finnish teacher evaluation system offers potentially constructive insight for U.S. education policy.

This chapter explains the research findings of this qualitative study in detail. It contains the research questions, information on the study sample, and the emergent themes in the data. The Finnish perspective is presented first, followed by the New York State perspective. A summary is also provided at the end of the chapter.

Sample

As part of comparing New York State and Finland, a sample of six educators from New York State and five educators from Finland were interviewed. Each group consisted of one or two administrators and four teachers, all of which taught or supervised at the secondary level. In the New York sample of educators, there were three men and three women. Additionally, five were White and one was Black. Educators were recruited from several regions across New York State, and members of this sample were from different regions of the state, such as Syracuse, Long Island, and the Capital Region. In the Finnish sample of educators, three were women and two were men.

Additionally, all five were White. Educators were recruited from several regions across Finland, and members of this sample were from different regions of the country such as Helsinki and Espoo. Participants were interviewed on the topic of teacher evaluations and supervisor feedback.

Also, the educators I interviewed in New York State were people I had never personally met before. In addition, many of the Finnish teachers I interviewed I had never met before, and the ones who I had met during my trip, I had only spoken to very briefly. It was beneficial to have been previously introduced to the Finnish educators in this case, because they seemed to feel more at ease with me as an interviewer from another country. Meeting some of the Finnish educators in advance was also planned for in my study methodology, as part of preparing for the logistical challenges that are present at times in recruiting participants who live in other countries.

Foremost Analysis Results

Five distinct themes emerged from the research data. The five major themes identified from the results of this study included:

1. Teachers appreciate feedback from supervisors whom they have a relationship with, and personally trust and respect.

2. Teachers and supervisors feel the use of high-stakes standardized test scores as a means to evaluate teacher effectiveness is an unfair and inappropriate practice.

3. Teachers cited many sources of helpful or influential feedback in their professional growth and development, including sources outside of a formal evaluation process.

4. There is a cultural difference in the way teachers and the teaching profession are perceived in New York State versus Finland.

5. The majority of teachers are very good, and the problem of the "ineffective teacher" is

small.

The Finnish Perspective

In this section, the perspectives of five Finnish educators are explored and analyzed. The information is organized according to the five emergent themes.

Theme 1: Teachers appreciate feedback from supervisors whom they have a relationship with, and personally trust and respect.

In Finland, there is a tremendous amount of trust in teachers and the feedback teachers receive is mostly positive. No one is "looking over their shoulders" at their teaching. If a parent or guardian expresses concern over a teacher, there is an open, two-way dialogue between the teacher and the supervisor. The teacher considers what the supervisor is saying, and the supervisor is open to the thoughts of the teacher. Problems are addressed through conversation, including times where the supervisor has to address something in a more direct, or "straight" manner.

> Like, between the lines, feedback from students. Also, boss and from colleagues, it can be straight, or not. But we have that kind of culture I feel that we are in a positive way. We care about what is happening and we are really sensitive to feel… sensitive to feel also to help each other… others at school as a colleague. So, we are talking all the time, how is it going? And I feel that we will have really easy help if we really need. But I hope that everybody is enough open to ask for help. I hope that teachers are so couraged. [Nora]

> I don't know others, but it's very little I get feedback from my supervisors. Because my class is going fine… After university, there is nobody who's looking at my teaching. Nobody's looking how I treat students. Not neither timing nor how I do presentations. So, it's on your own then. [Hans]

> I think my experience has been that they're very much like a conversation. That although we have some questions to kind of ponder about beforehand and such like, but it's not like an interview or an interrogation. It's very much of a discussion, where like, both sides … do you have anything which you would need more support with, or is there some change to school we can reorganize.

> So. I think… I felt very much that it was a discussion. That I could also bring up issues which I felt like that weren't new to me, but new to someone else. [Lily]

In the following two excerpts, Finnish Teachers share their reflections on times where their supervisors have provided them or their colleagues with feedback, including when they addressed things that they would like to see the teachers change:

> I think I have heard lately, two times that two different teachers had to go to the boss. And boss has said something really straight. And they have discussions. And I think those are little gossips, although I don't mean that we have a negative gossips going on. But I think it's it has been fair. There has been some teacher has had problem with parents or something. And she has talked about that with the boss, for example. So, it has been quite open case. And that's it. Yeah, I feel empathy. Okay? That's happened. Yeah. Such is life. Yeah. [Nora]

> And most of us, I mean, nobody's interested in how I'm doing my classes, this kind of side of it. Okay. You know what you're doing, okay? No worry about that. But then fill this form, make this requirement, do this inquiry. I mean, I do it happily. If I see why I'm doing it, I can see the results in future in my school or students. [Hans]

Because supervisors in Finland are in touch with the goings-on in the classrooms in their schools, there is little need for a more formal evaluation process that looks at test scores, rubrics, or numeric data to assess how teachers are performing. One teacher reflects on his relationship with his principal and how that impacts the governance of the school:

> I think without that type of relationship, then as the principal and for leadership then you would have to depend more on exams or some sort of feedback from the teacher to figure out what's going on with the students. [Kevin]

Overall, the relationship between the supervisor and the teacher emerged as being very important in the feedback exchange process. Problems are solved through two-way discussion where the teacher and the supervisor carefully consider one another's perspective, and come to agree on the solution. This way of solving problems in the schools is perceived as being a fair process by both teachers and supervisors.

Theme 2: Teachers and supervisors feel the use of high-stakes standardized test scores as a

means to evaluate teacher effectiveness is an unfair and inappropriate practice.

Finland does not have standardized tests. When asked about this practice and how it would be perceived in Finland, teachers in Finland gave direct and consistent answers, saying they did not want it. They also reiterated the system they have in Finland is highly trust-based, and that they feel teachers are the ones most qualified to assess their students' learning in their classes.

> I think we don't want that. No, no, no, no, not at all. We, we won't take it. No. The point is, I think, how students are learning, and learning is that we are focusing on. So that… that's enough to assess a teacher. As long as we will have good results in general, so there's no problem. [Nora]

> Not like it. [Laughter]. I don't know. Maybe, okay. If [standardized testing] comes it's fine, but maybe… it's totally new system. Totally new system. In Finland, the teachers, they are very independent. Very much independent…. Somehow, I think we have to trust when you are the professionals at school, That, we are working in very high ethics, we are prepared for the lessons, we are highly educated. So, all of these things. I mean, in the states also, we all are masters. We did studies for six years. And we did a lot of these studies. We love our work, and we always think about the best of our students. [Hans]

> [On standardized testing]. I think it's an outstanding idea there should be applied everywhere… NO. …. So, again, I do like it here in Finland where the principal is very hands-off. Principals are very respecting of the teacher's realm, or the classroom as a realm and their space. [Kevin]

> [On standardized testing]. I think it would go against everything we believe in. Because we kind of believe that everyone has an equity and everyone has an equal… should have an equal right to learn and to develop. And if we are punishing school teachers, to work in schools, where the students are lower achieving, then in schools where, students are high achieving, then it's not taking into account the improvement that has happened while they're there. … And the whole standardized testing thing is so problematic anyway, because you get the whole teaching to the test. [Lily]

One supervisor in Finland shared that he would be "devastated" if this practice came to Finland. Based on his experience, he estimates that 75% of students' performance is affected by unknown characteristics, and that the practice of evaluating teacher effectiveness based on this would be unfair and flawed.

> In a word, devastated. … If I have that class that I just mentioned to you, whose classroom performance is affected 75% by the unknown characteristics, how can I be expected to perform as well as a teacher who has a homogeneous group of students who get excellent results? [Ron]

Finnish educators pointed out that you really have no way of knowing what emotional state students may be in, prior to learning. There is no way to measure this, other than teachers' informal observations of their students, and their making note of any visible signs of emotional distress. But some students are also very good at hiding their emotions, and we really do not know what a student may be going through. If a students' emotions are not regulated prior to entering the classroom, let alone a high-stakes standardized test setting, this could have an impact on their learning and performance.

> Yeah. You've got to look at culture, you've got to look at things like socioeconomics, I mean, these youngsters are bringing all kinds of issues with them to school – did they sleep well last night, have they had a proper meal in the last 36 hours, has the boyfriend just finished with them, have their mum and dad been fighting? I mean, we do not know those, all we see is this polished version of society sitting in front of them. And I think that their characteristics, in my opinion are 75% affected by environment. [Ron]

> No, and that's the thing too is I see a lot of students who were like crying in the hallways, or they're talking on their phone, screaming or something and they're emotional, and then they're obviously not in a state of mind to retain information after this. So, I just think there's always that issue, and then the student walks into my classroom. I don't know if they were outside in the hallway crying, because I was in my classroom. Like so, maybe they cleaned up really well, pulled themselves together, and now they're there and I'm like, Hey, like, you're two minutes late, and I'm like going off on this kid for being two minutes late. I don't know what happened. But they're obviously not going to retain information. [Kevin]

Theme 3: Teachers cited many sources of helpful or influential feedback in their professional growth and development, including sources outside of a formal evaluation process.

Teachers cited many different sources of helpful or influential feedback in their professional growth and development. For example, even though teachers do not typically observe each other's lessons directly, teachers in Finland find interactions with their colleagues to be very beneficial. Many teachers say they wish they had more opportunities to work with their colleagues. Some specifically mention that feedback from their colleagues helps them, and that more time with

their colleagues and opportunities for peer observation and feedback would be helpful.

> Colleagues, I guess don't have that much. I mean, they don't really give me feedback on my teaching, but they don't really come to our lessons. We don't really visit each other's lessons that much. [Lily]

> I think it might be good… that could be like…. if the colleagues could visit other classes maybe once a year. And kind of small piece of paper, the presentation, the order, how did you do, and like that kind of feedback in a friendly, but honest way. And could we then share that together. … Because it's very rare, you'll get straight frank feedback. But if it comes from here (motions like someone is watching you from above), that somebody's watching you like you know the writer George Orwell. So, you know, the writer, formidable novel [Animal Farm] … Yeah. But if that's kind of same level. I mean, your colleagues, they know you. They can be frank. I don't mind. It would help me to grow. [Hans]

One Finnish teacher expresses how she feels supported by her supervisors, and that they provide teachers with positive feedback that makes them grow.

> Definitely. Yeah. I feel that the bosses are on our side. That they want to encourage us to grow. They offer more educational courses, they say about them, sometimes. I think they give us teachers the positive energy to grow. [Nora]

Additionally, one supervisor in Finland spoke to the importance of having positive relationships with the teachers he supervises, and how positive interactions help his teachers grow.

> I'm sure both of us have been there, where somebody has said to us that was poor, that was not very good. You need to do that better. And we're experienced enough to probably accept that, but younger teachers may find it very, very hard if they're being criticized, and then we drive them out of the profession, when all they needed for a short period of time was their hand-holding. They needed their confidence boosted, because aren't we better people when people are saying positive things about us? [Ron]

Other teachers and supervisors in Finland cited reading as being helpful in their professional growth and development, as well as reflecting on feedback they have received from students.

> I like to read a lot of pedagogy books and develop my, my own system with my students. So, students are my teachers. That's fine. Also, of course, those books and theory background, but I love that feeling that I'm quite free as a teacher to develop pedagogy with children and students.

So, all too less, I speak with my methods with my nearest colleagues, but also I'm a teacher trainer. So sometimes I have a chance to share my ideas that delights me a lot. … There I meet really nice and creative and talented other teachers, so I can say also colleagues are my teachers. But at school, I feel that we have never time to really speak about our ideas with our nearest colleagues. [Nora]

And also, I try to evaluate my teaching also. I do every year for older students also, that they evaluate my teaching. Do I speak slowly enough, or clearly enough, or my notes - are they in order, or exams, or working during the classes and they can put some comments. [Hans]

A few teachers talked about how strong relationships with their principals, where they can comfortably participate in two-way dialogue, were a contributing factor to their professional growth and development.

So, having that kind of relationship with the principal was super critical. Where I think other teachers… even though [the principal] is very open-door policy, all the staff is…. Finns… it's difficult for them to step over that threshold. They really only want to go to them if it's something critical. … Otherwise, I'm a pretty stubborn loaner when it comes to my development, I have to do the trial-and-error thing a lot in my life. While people watch from a distance and they're like dude you're crashing and burning, you should've talked to me. I'm like I know. … They don't know my curriculum that I teach, because I'm the only business studies teacher. So, it's not like I can talk to other business studies teachers about my curriculum. It's my own department within humanities. So, I'm an island. [Kevin]

One educator talked about the reasons why feedback from supervisors is not always a contributing factor to their professional growth and development.

I was a PE and a maths teacher then and I had a maths lesson and my supervisor from the university came in. The only criticism he could find was that I said "uhhh" or "umm" sixteen times. I thought to myself, what kind of feedback is that. [Ron]

One teacher spoke on professional development, and on what she feels are essential elements that make for a meaningful professional development experience.

I think the Finnish solution to that was like, you used to, when you trained to be a teacher at university, you used to get a grade for your teacher practice, like the practical part where you would teach real students in like the teaching schools, you would get a grade for how good you are teaching. And then that would actually affect like your job application. Thank God they got rid

of that because of course, that was completely skewed, it depended on who your supervisor was and what kind of style of teaching they liked. [Lily]

Theme 4 – There are cultural differences in the way that teachers and the teaching profession are perceived in New York State and Finland.

In Finland, teachers do not have a formal evaluation system per se, but rather teachers are provided feedback through conversations with their supervisors and colleagues. This feedback comes infrequently, due to the fact that teachers are given a large degree of autonomy and are trusted by the public in Finland. Several Finnish educators mentioned in their interviews that they feel the reason for this public trust is because the rigorous schooling system teachers complete to become teachers is held in high esteem in Finnish culture.

> When you said that, I was thinking that we have such a good quality at our teacher training. Teacher education is so good that every teacher knows that every teacher has learned through those standards. So, I think we are thinking that way. We all already passed that school and we are good teachers. Of course, we have differences, but not that much that it's a number that we have to talk about. I cannot say for example, at my school that they are better or worse teachers because I think we are quite in the same level. We have different interests. And I think also teacher education… we can be personal, we can have our strengths there. So, we have our culture has already accepted that teachers are different and we have different strengths. And after that standard, standard education and school teacher education of course, then we trust…. trust others that we are almost at the same level. We talk about same things. [Nora]

> Somehow, I think we have to trust when you are the professionals at school. That we are working in very high ethics, we are prepared for the lessons, we are highly educated. So, all of these things. I mean, in the states also, we all are masters. We did studies for six years. And we did a lot of these studies. We love our work, and we always think about the best of our students. [Hans]

> I mean it's a case because the system, in theory, is so good. And because the teachers are so qualified, we automatically have an expectation that because of that, they know what they're doing. And then, if not, then it's a very guiding principal kind of idea that we're there to help, not to chop their legs off. I mean, we're there to provide the guidance and the support, maybe from our own experiences like I mentioned earlier, but more often than not, it's a "let's make you better than you are" kind of approach, rather than "let's focus on the problems that are lying there." Because that's devastating to hear. [Ron]

In the second quote above, Hans mentioned that teachers in the states have to go through the same schooling as Finnish teachers. This is true. In both New York State and Finland, teachers must have a master's degree in order to be qualified to teach. However, in Finnish culture, obtaining this master's degree results in a high degree of trust and respect. In New York State, teachers are not treated with the same respect and culture, even though they had to go through the same rigorous schooling as Finnish teachers.

One Finnish teacher expressed that some elements of accountability would be a welcome addition in the Finnish system, sometimes. This acknowledges that the Finnish system lacks the ability to remedy ineffective teaching practices at times. However, the teacher also admits that while the added accountability would be helpful, it would not be worth the heavy stress, if that were to be a natural unintended consequence of such an addition.

> The oversight teachers have here in Finland is the rigorous schooling system. Again, the teachers are hired off university without real world experience. There's very little oversight of the teachers once they're a teacher, and then here they do a one year work contract, and then if they like you they sign up for a permanent contract. So, after two years, just have two lucky years of being a good teacher, and then you're a permanent teacher, and with the unions here you can't get rid of that teacher. So, there should kind of maybe be some oversight. Some flexibilities with being able to get rid of teachers and fire them, but at the same time, so maybe Finland is too liberal like that… too protective with the unions and the employees with them. But at the same time, I don't think we should have what's in the states, with this heavy blanket of stress. [Kevin]

Lily, another Finnish teacher, offers her perspective on the cultural aspects of teacher evaluation systems in Finland, versus New York State.

> It seems like in Finland teachers are very trusted. And that is just this whole underlying phenomenon that really drives a lot of what works in Finland, and in the United States teachers are not trusted. They're, you know, doubted and questioned, and just everyone always kind of assumes a negative view of what might be going on in the classroom. And it drives the policy, because like then we have these lengthy evaluation systems, and we have these test scores, and are our teachers doing their jobs? And it's very negative sometimes. [Lily]

> Hopefully here I think that hasn't been a big impact, because we have such strong trust in teachers. That's why the whole organizational system is so light, because we trust that teachers want to do their job well, and they don't need to be observed to do it well. [Lily]

> I think the trend here has been going a little bit in that direction, in the way that the national curriculum is built, etc.. Like, year by year, there's a little more of that, can we trust the teacher to do this correctly? I mean, our curriculum is still nothing compared to a lot of countries' curriculums that almost have lesson plans, ours are very general, but the direction more towards like regulating what is done in which year, in which order, and so on. So, there is a little bit more of that. [Lily]

Overall, Finnish educators appear to be satisfied with the way they are perceived in society.

Theme 5: The majority of teachers are very good, and the problem of the "ineffective teacher" is small.

Many teachers and supervisors in Finland had much to share on the topic of ineffective teachers. There was a consensus across the participants in this study that there are very few ineffective teachers. In addition, there are some effective teachers doing some ineffective things, where interventions could be made to address and mitigate the ineffective behaviors occurring.

> … the issue is more commonly things like class management, that it's not working so that the students can actually concentrate because there's too much disruption. Or, or that, that it's too complicated for the students also some cases that the teachers somehow haven't understood what level the students really are starting at and haven't made it simple enough to begin with. [Lily]

Teachers and supervisors in Finland reflected in their responses that the teaching profession is viewed in high esteem. Additionally, educators in Finland commented that there are very few ineffective teachers.

> When you said that, I was thinking that we have such a good quality at our teacher training. Teacher Education is so good that every teacher knows that every teacher has learned through those standards. So, I think we are thinking that way. We all already passed that school and we are good teachers. Of course, we have differences, but not that much that it's a number that we have to talk about. I cannot say for example, at my school that they are better or worse teachers because I think we are quite in the same level. We have different interests. And I think also teacher

> education… we can be personal, we can have our strengths there. So, we have our culture has already accepted that teachers are different and we have different strengths. And after that standard, standard education and school teacher education of course, then we trust…. trust others that we are almost at the same level. We talk about same things. [Nora]

> Maybe. Some. Some. I really have to think about it. I can think about that whole career of my 21 years maybe one. Not my close colleague, but she… I think she had some mental problems. [Hans]

One supervisor in Finland spoke about the impact of an ineffective teacher. While there may be very few ineffective teachers, the impact of an ineffective teacher on a student's future trajectory can be profound.

> So, if students go into classes and they've really disliked the teacher, disliked the teaching approach, they were unsuccessful, they were labeled as not being very good. They'll remember that. They'll remember that for many, many years to come. I think that the impact of an ineffective teacher is not just at that moment in time, in space. I really believe that the influence that we have on molding and creating youngsters is as… I think we often underestimate that. Things that they remember, the things that they want to forget. School in general can change a youngsters life inevitably and indefinitely. What their experiences are in school will then set the course. And believe me, they will remember. [Ron]

In addition, one teacher in Finland spoke about how Finnish teachers address concerns with their colleagues, through conversation

> In our humanitarian department, we politically and kindly tell the person that we have expectations and try to be clear on expectations in a nice way. And let them know that that bar needs to be set at, at least that level. And if it's not, then we all need to then step up. So, we all have families and other things going on. Let's try to make sure that bar is set there so that we're all equally pitching in. Lay it out clear because, they don't know unless you tell them. [Kevin]

The New York State Perspective

In this section, the perspectives of six New York State educators are explored and analyzed. The information is organized according to the five emergent themes. It is also compared with the data presented from the Finnish educators.

Theme 1: Teachers appreciate feedback from supervisors whom they have a relationship with, and personally trust and respect.

In speaking with New York State teachers about feedback that was meaningful and influential in their development as teachers, a clear theme emerged regarding who the feedback was coming from. Many teachers shared stories or examples of times where a supervisor provided feedback that had a positive impact on them personally, and the teachers took that feedback into consideration for their own improvement. Conversely, many teachers also shared examples where the supervisor provided feedback that teachers thought was misguided, out of touch, or did not align with their own perceptions of their classrooms, and thus did not take the feedback into careful consideration. Overall, teachers valued and considered feedback that was provided by someone whom they personally respected.

Some of the qualities of the respected supervisors were that they delivered feedback in a kind and/or positive way and that they already had an established working relationship with the teacher. Respected supervisors' communication of the feedback reflected that they had an awareness of the context of the classroom, and could empathize with what the teacher was experiencing with his or her students.

> And to have, you know, someone that is kind enough, I guess is a better way to say it. They weren't, like, telling me what to do. There was no like, you know, you're not good enough feeling in that. [Lana]

> That you fill that [same information in] for me and Joe Schmo down the hall. That's not specific to me. So, I guess I look at… I'll take the feedback, but just make sure it's feedback for me. [Leah]

> So just that concept of wanting to relate to those that she [the principal] works with and wanting to be a part of their classrooms, I think is has really helped a lot and it's… I think it's very… it's a very positive… kind of starts your day really positive. And I think we need that. And a lot of teachers need it, and the kids need it too. And that's most helpful. [Lana]

And my principal I think did a really nice job with that. Not that she was perfect in any way, because I don't think anybody can. This year was very hard, very, very hard. But enough so that the way she put it, the way she would say things… it made you feel like she appreciated what you were doing. And just was kind of showing you a different way to do it. That's huge. It really is. Having that relationship with your colleagues is a big deal. I'm trying to think… if you weren't super motivated to try and fix whatever it was that was a problem…. I just I don't think you would. [Lana]

I view them [APPR observations] as subjective and I think depending on who actually does it… like again, if you don't know me as a teacher, and you're just some random person walking in… what you have to say about me, I don't put too much weight in. You know. Whereas, that time where the observer came in and he was like, "I really think you need to be teaching more," I was devastated. Because like, I respected him so much. If somebody else told me that I'd be like, whatever, but like, he knew me, and that was devastating to me. [Leah]

Frequent, natural feedback where the supervisor was in touch with the everyday climate and conditions in the classroom was also impactful.

He [the principal] would pop in the room and I was like do you need something? He's like, "I do this a lot." But I had come from an environment where, you only saw the administrator when there was an issue, where they were doing some type of an observation, you know, so that was different. So, like, he was just constantly swinging in, checking in, not even like giving feedback, but like, just are you good? You know what I mean? So, it's like it made it a lot easier when he came in for those formal [evaluations] in here too. [Leah]

Like I said that guy… when he did that, he wrote God bless the four of you. I was… it just validated that like, I'm trying. It might not look great or whatever. But that was all I needed. I was like, I'll ride or die for you all day long. You know, I mean, it doesn't take a lot. But that stuff matters, you know? [Leah]

Feedback that lacked these qualities was off-putting. Rather than sparking reflection in the teacher, the focus became a lack of awareness on the part of the supervisor.

So, [the lesson] did end up being a little chaotic. And he was kind of like, yeah, it was a little chaotic. And I was just like, yup, alright. I kind of shut down. It's hard! Well, then you [the supervisor] should have at least given me a heads up you were coming, and I could have done something that would have made it they didn't have to leave the classroom. I also should have had the computers, because that's what you promised me when you hired me. And I just wasn't sure when he walked in the room. I wasn't sure whether this was formal or informal, and he had only been in there… this was the first time I think he'd actually been in there. … So, I mean, there's just, there's ways of talking to people and doing things. You don't have to hurt them. [Lana]

Similarly, one teacher expressed concern over receiving feedback from a supervisor who lacked credibility.

> Like, you know, when you have people sitting there trying to critique you who have never had to do anything COVID-related... [Leah]

Teachers also shared that conversations and narrative written feedback were more impactful and motivating delivery methods than providing numeric feedback, such as from a rubric.

> I think too, the feedback needs to be more than numbers. You need to have some things, some narratives, you need to have some conversations. But I also think it's useless to have like generic stuff on there, like you know, should be pre-set, you know, should have standards on the board. Like what does that... you fill that in for me, and Joe Schmo down the hall. That's not specific to me. So, I guess I look at... I'll take the feedback, but just make sure it's feedback for me. [Leah]

> Yeah! Like especially, if you're going to take the time and actually write something. Like clicking a number, I don't care. If like, all you have time for is, I'm just the number to you, you know, it's kind of like, I'm not really... that's fine. But if there are some, I mean, positive feedback too, so I mean, that's great. But like if there's times that there's a recommendation, I 100% take that into consideration, because maybe I haven't thought about it, or maybe I've tried it and those are the times I go back and I'm like, I've tried this and let me tell you how it didn't work. But like I hear you, I tried it, it doesn't work right now, it might work later. So, I think... I think it helps motivate me. [Leah]

Theme 2: Teachers and supervisors feel the use of high-stakes standardized test scores as a means to evaluate teacher effectiveness is an unfair and inappropriate practice.

It was commonly agreed upon by both teachers and supervisors, as well as educators in both New York State and Finland, that the use of high-stakes standardized test scores as a means to evaluate teacher effectiveness was an unfair, unhelpful practice that is based on erroneous assumptions. Many reasons and explanations were provided for this stance, but in general, educators feel that there are many factors that lie outside of the control of the teacher that impact

students' performance on such tests, and it is flawed and unfair to assume that students' successes or failures are strongly related to, or even caused by, students' learning in the classroom from their teachers.

One teacher from New York State captured both the flawed nature as well as the "unfairness" sentiment of the APPR evaluation system with a clever analogy.

> I mean, it's just like a dentist. Are you really going to rate a dentist based on how many of their patients' teeth are doing well? I mean, you can't do that. It doesn't work that way. But I don't know that I would have an answer to the best way to evaluate a teacher. I think, just like you do your students, getting to know the teachers and do what my principal does and really find a way to make that connection. These little team huddles, being able to work one-on-one with the people that are doing the best they can to work with these kids. But, I don't think that's enough. I don't think observations from administrators is enough. We need to we need the training. We need the professional development. It's a need. [Lana]

One supervisor from New York State shares his thoughts on the practice of using high-stakes standardized test scores as a means to evaluate teacher effectiveness.

> I think it's so misguided, it's horrendous. And, you know, that's like me…. I liken it to taking a scale, and using a scale to determine somebody's height. Two different tools. They're both important tools. But if I'm trying to measure someone's height, give me a measuring tape, not a bathroom scale. So, do those tests play a role in teaching and learning? Of course, they do, but they're not designed as a way of evaluating teacher effectiveness. But, that's how New York State decided to use them, along with many other states. So, yeah, I think it's totally misguided. [Stephen]

One teacher brought up that if high scores on standardized tests are looked upon favorably in the evaluation process, then how would it be possible for teachers of students who struggle on tests, such as special education teachers, to score well on their teacher evaluations?

> [Regarding state test scores] Well, like, I think it's ridiculous. Because again, like, who wants to be a special ed teacher? If you start to do stuff like this? [Leah]

One teacher in New York State spoke more broadly about the unfairness of the tests, and the political nature of the practice. He talked about the tests being biased, and at times impossible

for students who grew up in conditions where they cannot relate to the test material. Now, the culture has become such that educators are resorting to extreme practices, such as cheating, to obtain satisfactory scores in an effort to save their jobs.

> From the beginning, I was against [using standardized test scores as a means of evaluating teacher effectiveness]. I was like you know, again, it's a political effect of our education system. I feel like it creates, or it created a lot of issues, a lot of problems in the education field. … and I heard some stuff of the school district around the country where now teachers are doing everything to get high scores for their kids, even if they have to cheat. And that's really… some schools got in trouble for that. So, it's almost like we're getting away of teaching, and then we're focusing on now saving our jobs. We're teaching to the test, even though that doesn't really reflect reality in real life. And the way we teach is different from the way those tests are administered. The way even we ask question on our quizzes on our weekly test is totally different from the way the state test is. … The state test is so biased. You know, they ask questions to students who… let's say a student from the suburb, who parents have enough money to put them into a hockey league. They go swimming during the winter. City kids don't even go swimming during the summer. Forget about hockey, because those are expensive schools, forget about lacrosse. [Yael]

Many teachers brought up different examples of factors that influence test scores that lie outside of the control of the teacher. Some examples include resource insecurity, such as food, clothing, or shelter.

> I don't think it's an effective use of data, because again, we don't know what the kids are coming in with… the baggage… particularly after the last couple years… And is it fair to put it on teachers? No, not necessarily. Because what happens the night before an exam at home? Or you've got kids I mean, you know this, you've got kids who are just worried about finding food, or supplies, or wearing the same shirt to school four days in a row. And then we look at the pressure of putting that on, and there are only so many things that we as human beings, as teachers, can do to ensure how their success will be on some type of test. [Miranda]

Another teacher pointed out that students' attendance in school, suspension from school, or other unique family or medical reasons that cause students to miss years of school, and thus have gaps in their prior learning, may be influencing test performance.

> As a classroom teacher, people always think that you have the power, you have the control, which is in some sense is true. You're the classroom teacher. You're the one who needs to make sure that they understand that the message gets through. They spend more time with you than

anybody else. But at the same time, so many things are most of the time are in your way to stop those kinds of things. I mean, attendance. If kids are not coming to school, that's out of your control. Suspension, if the kids are suspended, even if you provide work, if it's different than the kids being in school, then doing their work or showing you their packets at home. If the students also miss months of school, or years of school, or if the student was in a special program in a school and it was discontinued, even though the students was not ready, or students are already below grade level before he or she gets to you. Really how much can you do, especially if you don't have the resources to help the student. And sometimes also, the curriculum is not… is doesn't really reflect what you need to help the students. [Yael]

It's not just about what happens in my classroom. And great, they can write a solid essay, or they can read, but it's so much more the little things, the relationships, the stressors that these kids are on, that you or I, we didn't have to deal with. So many things. Social media. Just the things that I think a lot of people don't recognize, even though we see them every single day, even though they're in our classrooms, that… how difficult it is for the kids to shut their brains down. [Miranda]

In summary, there are a variety of factors teachers cite as being out of their control as teachers, yet impacting their students' test performance. Because of this, they feel assessing their quality as teachers based on this data is an unfair and flawed practice. Overall, the use of this practice goes against the idea of trust and respect for teachers, which was shown in Theme 1 above to be so important in the feedback process.

Theme 3: Teachers cited many sources of helpful or influential feedback in their professional growth and development, even outside of a formal evaluation process.

Similar to teachers in Finland, teachers in New York State also cited many different sources of helpful or influential feedback in their professional growth and development, even sources that lie outside of a formal observation process such as APPR. Even though peer observation is not a common practice, teachers in New York State and Finland both expressed that they find their interactions with their colleagues to be very beneficial. Teachers in New York State also expressed that they wished they had more opportunities to work with their colleagues. Some specifically mention that feedback from their colleagues helps them, and that more time with their colleagues

and opportunities for peer observation and feedback would be helpful.

> I would say being able to work once again with colleagues. And, my principal this year has really put a lot of effort into trying to work with the teachers to help them in different ways. They, you know, we have meetings as a team… they'll do meetings, she calls them huddles, which I think is kind of funny. … (On what she would like more of) So, I guess more of a connection with my colleagues. I'd really like more of. Probably every teacher would say the same thing. [Lana] [Lana]

> So, I would say pushing myself but also really working with my colleagues. I really bounce ideas off of them and vice versa. Technology, learning about technology from some of the newer teachers has been very, very helpful because you know, we all get set in our ways. But that has been really, really inspirational, I think, having them help me with certain things like that. [Miranda]

> I will say my experiences, and then also my self-motivation for finding things that I need outside of what the school provides me. … Or sometimes I mean, we talk among colleagues, even though colleagues is not there when you get observed, but you can always go and ask them, you know, for feedback. [Yael]

Similar to the supervisor in Finland, one supervisors in New York State spoke to the power and importance of two-way discussions with the teachers he supervises about best practices. The supervisor in New York State was also mindful and reflective of *how* he goes about having good conversations with staff, so that the content of the conversations is impactful.

> I think it's probably the collegial conversations, learning from other supervisors, kind of engaging in good discussions about best practices. Learning from experiences, good and bad. And then challenging yourself to grow. So, I think having that continual perspective to grow as a leader, learn more, learn from other professionals, continue to grow in your own craft, and then applying them. So, I think that's been a driving force for me, as well, at least that constant. You never really achieved the pinnacle of "I'm there." [Brian]

> You have a continuum of people. Some would do it regardless of it, because that's the way they are, where other people need that collegial push. You know, and then providing good feedback, you know, and I think that's one of the challenges also as well, is where do people need to grow and how do you identify it and, you know, how do we do that in a way that's getting the greatest good. There's only so much time and effort that people are willing and able to shift sometimes. So how do you maximize on what the most appropriate feedback is to help them develop in their craft to help their students. [Brian]

Similar to educators in Finland, some teachers and supervisors in New York State cited reading as being helpful in their professional growth and development.

> I do a lot of reflection. I do a lot of reading. I read both school-related stuff, I read a lot of history. And I think as you read and learn about how leaders and how they navigated difficult situations, I think that can be very informing. I think my conversations with parents and families have been very helpful. You know, work with union leaders, you know, what is it that I'm communicating, what are people hearing, in terms of what I'm communicating? And that's been very helpful as well. So, just making sure I'm getting some feedback loop and listen to it. That's all been helpful. [Stephen]

One teacher talked about how her strong relationships with her principal, where she could comfortably participate in two-way dialogue, was a contributing factor in her professional growth and development.

> [The principal] would, you know, be a part of the lesson. She would privately chat with me, but with the kids she was more just another person there, like a TA type thing….Which is another way of observing… you know… it's another way of using that process. I think it's a better way, because not only were you, like I said it's more of a community, you're involving the teacher, your students, the TA that's already in there, and then you're now involving an administrator. So, it is more of a conversation, instead of like a test. [Lana]

A few New York State teachers talked about the reasons why feedback from supervisors is not always a contributing factor to their professional growth and development.

> I think [teachers are] more scared. Does that make sense? They have a fear of these observations, which I'm always, like, I don't understand how you can be afraid, you know what you're doing is well done. And if they give you a negative comment, just use it to do better. I don't know…. but people, I mean it's a normal human being reaction to anything like that. Nobody wants to be told they're not doing something right. Nobody wants to be told they have to fix something… but we have to. [Lana]

> So, it's really, it's almost like sometimes I feel like supervisors just tell you, we just give you the information. And hoping that you understand that they don't really have the answer. You know, they just give information to you. And then now you've got to figure it out. This really, it's a tough, you know, and me as a classroom teacher, you know, that you're the one in the end who's going to deal with the students in the classroom. You know, because the supervisor, maybe you see him once a while, but you are 24 hours a day with the students. And now it falls on you to figure out with a special ed teacher, with an ENL teacher, how are you going to provide instruction, or to help these kids as, you know, the best of your abilities. So, it's a tough area. [Yael]

One teacher spoke on professional development, and on what she feels are essential elements that make for a meaningful professional development experience.

> I think people like to do professional development, if they see a value in what they're doing, and if someone takes that time to value the people in the room. Professional development that's forced on you, that it is not for everyone. It doesn't work, and it's boring, and you barely get anything out of it, if you get anything. So, once again that's another issue. I think more of our evaluation needs to go into providing good professional development [Lana]

Theme 4 – There are cultural differences in the way that teachers and the teaching profession are perceived in New York State and Finland.

The general culture and public perception of the teaching profession in New York State is different than it is in Finland. In New York State, the public can be distrustful of teachers at times, and express skepticism over teachers working to their best ability. This reflects the larger market accountability model that exists in the United States, including New York State. Teacher evaluation systems organized around market accountability involve clear signals to the "market" about the effectiveness of schools and teachers, and the use of student achievement test scores are often used to provide this signal (Williams & Engel, 2013). In Finland, there is more trust and professional accountability, and thus less need for educators to justify their actions.

New York State teachers expressed some concerns over the current APPR teacher evaluation system, particularly in the area of fairness. Teachers feel certain systems need to be in place to address pieces of their evaluations that lie outside of their control, such as student attendance. This speaks to the culture of distrust of teachers in New York State. This also subtly reflects a culture of market accountability where teachers are expected to produce results. These results could be in the form of achievement on high-stakes standardized tests, or a measurable growth score in students' academic improvement for the year. One New York State teacher shares

his thoughts on what needs to change to improve the fairness in teacher evaluation system in New York State.

> I think a lot of things needs to change. I also like the accountability on teachers, but maybe that needs to be designed in a different way. … For me to be confident the evaluation is going to work is that a things have to happen. You know, make sure kids are in school. Make sure parents are involved. Make sure parents know what's going on with their school, make sure that when we have back to school night, when we get the notice, it gets to parents ahead of time… you know, heads up, so they can come to the meetings. Make sure the resources are there for the teachers. Make sure the curriculum matches whatever tests they're giving us. You know, make sure the kids have… we don't discontinue the kids [meaning, declassify students from receiving needed services] [Yael]

Theme 5: The majority of teachers are very good, and the problem of the "ineffective teacher" is small.

Many New York State educators had much to share on the topic of ineffective teachers. There was a consensus across the participants in this study, New York State and Finland alike, that there are very few ineffective teachers. In addition, there are some effective teachers doing some ineffective things, where interventions could be made to address and mitigate the ineffective behaviors occurring.

The way ineffective teachers are identified an addressed in Finland is different than in the United States (New York State). This is because the United States (New York State) has a culture of implementing market accountability practices, versus Finland which has a culture trust in teachers and thus operates on professional accountability practices. Finnish supervisors only learn about concerns in teachers' classrooms through discussions with colleagues, families, and students. Finnish supervisors addresses the concerns through conversations with the teacher about whom the concern was raised. Similarly, New York State (NYS) supervisors also identifies concerns through both feedback from stakeholders, but they also have additional insight on teacher

effectiveness through the APPR evaluation process. NYS supervisors address concerns through conversations just as Finnish supervisors do, however they have the additional option of writing Teacher Improvement Plans (TIPs) to outline a plan for remediating the concern, in the case of untenured teachers.

Unfortunately, the APPR evaluation process in New York State results in all teachers being assigned a designation of Highly Effective, Effective, Developing, or Ineffective. As mentioned in chapter one, this is based on a state-approved rubric, consisting of both observations as well as state standardized test scores. Thus, teachers may receive a designation of Ineffective, when really their teaching practices in the classroom are quite excellent. Conversely, it may result in a teacher receiving a designation of Highly Effective, when the teacher may actually have poor relationships with his or her students, or exhibit other ineffective behaviors in the classroom, or professionally. As mentioned in Theme 4, New York State educators feel that the use of high stakes standardized test scores as a means to evaluate teacher effectiveness is an unfair and misguided practice. They feel different approaches are needed to identify and adequately address ineffective teaching practices in the classroom.

Teachers and supervisors in New York State cited many examples of ineffective behaviors that may be happening, such as someone who is not trying to help his or her students, someone who has poor work attendance, or someone who is not really teaching, or someone whose teaching methods have not evolved with the times. It was also brought up several times that there are very few ineffective teachers, and that most teachers are of high quality.

> I think the majority of educators and schools across New York in particular are solid teachers, good teachers. You know, there's definitely probably the bell curve doesn't get applied as much from when you look at how the evaluation system identifies where people are in the continuum, or what they could do different, or who's the 10% that needs to get a little better or do things a little differently or change the pedagogy or catch up with the times. [Brian]

> Most of the people I have worked with have been incredible. Like they just worked really hard. And I don't know if it's a central New York thing or not, but this year, um, you know, public schools kind of get a bad rap. They have done, at least my colleagues have done so much to try and help these kids, and I would say if someone was ineffective, it would be somebody who wasn't trying to help their students at this time. [Lana]

> I've worked with a lot of teachers. Ineffective? No. Like, maybe needed more support and didn't get it? Probably. That's probably due to the fact that, again, we see these people [administrators] twice a year out of 180 days. So, you know, or my favorite is, you know, administration could hear from a parent, and now it's a problem. You know, it's been going on, but now there's a parent complaint, there's a concern, and now they want to address it. [Leah]

> And again, ineffective to me, like, I don't know, like you're out 900 days, you're not communicating with people. I don't even care if kids or parents like you. I think that's ridiculous, but it's like, are you... like, I mean... ineffective would just be, I don't know... I guess it's because I haven't seen it. I'm assuming ineffective would be like you're not there, or you're not really teaching, or like you're just targeting kids. [Leah]

One New York State teacher also commented that effective teachers make connections with students, and try to help them.

> I think, makes a connection with the kids that they might not have... they may not ever have had. So, I think the more that you can do that, then the more successful your relationship can be with the kids. [Miranda]

A New York State supervisor agreed with this sentiment, and elaborated by stressing the importance of a teacher's ability to recognize students' gaps in learning, and to reach out to other professionals in the school system when a student has a need that lies outside the scope of the teacher.

> Recognizing in the gaps of learning, we speak to the need to a different pedagogical approach. So that's really important. And some kids may need a different delivery than others. So, I think it's important for teachers to be able to recognize if students are having gaps in their learning, and to be able to address it somehow through a different approach or some scaffolding, reinforcement, whatever it might be. I think that teachers struggle clearly, they can't control the home life, which has an important impact on student learning and student retention. And that's where the whole school... but what teachers can do if they're noticing problems, and this is this is something that we really stressed in my previous district, let somebody know, let guidance know, so at least, you know, we don't wait to see a great drop or something like that. [Stephen]

Also, sometimes teachers in New York State are designated as being 'Ineffective' on the APPR rubric, based on their students' scores, even though high-quality teaching is actually happening in the classroom. A New York State supervisor and a New York State teacher commented on the use of this practice.

> There's a wholeness to an organization, and to take the wholeness of a school, which is not made up of individual silos, but rather a community and start picking out little pieces, I think is misguided and loses the sight of the school as an organism, right, living and breathing. So, are there ineffective? Yes. Is that the way to address it? No. And, you know, we had a case in one of my, in my previous school where one of the teachers who was outstanding - multiple times PTO Teacher of the Year, rave reviews from students - was determined ineffective because she was teaching kids who most struggled in mathematics, and their scores didn't show enough growth, according to this mystery formula the distinct reviews. And the sense that that was the most ineffective teacher in our school? Everyone was saying, you've got to be crazy. You know, I wrote an article about that that was published. It's just ridiculous. So, there are ineffective folks, that's what good evaluation systems help identify and support. And I would argue that the use of student test scores that identify them is misguided. [Stephen]

> I feel like the idea of an ineffective teacher, it could be so many things. And that it may not have one blessed thing to do with a test score. It could have to do with relationships in a classroom. I think the most effective teachers are probably the ones who command respect, but because the kids want to give respect rather than feel they have to. [Miranda]

Summary

This chapter presented the analysis of data collected from interviews of 12 study participants. Six participants were from New York State, and six were from Finland. Within each group of six, two were supervisors and four were teachers. Gathering the perspectives from educators from both New York State and Finland made a comparative study possible, and provided the necessary data in order to code and extract themes. Five foremost themes were identified and discussed, as part of comparing teacher evaluation systems in New York State and Finland. Sources of feedback that serve to motivate teachers to grow and develop were also explored.

The primary research questions for this study were: How do teacher evaluation systems in

New York State compare with those in Finland at both the structural and cultural level? How do teachers in New York State and Finland view the contribution of teacher evaluations to their motivation and professional growth? What are the policy implications resulting from the above comparison? These three research questions were used to develop the interview questions in the study instruments that were used to interview teachers and supervisors.

The supplemental research questions for this study were: How satisfied are teachers and supervisors with teacher evaluation practices in New York State versus Finland? To what degree do teacher evaluation practices contribute to professional growth in New York State versus Finland? These supplemental research questions are more focused on specific topics.

The findings for each primary research question and supplemental research question will be discussed in further detail in chapter five. Chapter five will organize information by research question, and will connect back to the theoretical and conceptual frameworks for this study. Study limitations, as well as implications for future research will also be discussed.

CHAPTER 5: DISCUSSION

Introduction

The purpose of this study is to understand potential alternatives to APPR as practiced in New York State, such as the case of Finland. Additionally, this study seeks to understand different types of supervisor feedback and their effect on teacher motivation. In chapter four, qualitative data collected from six New York State educators and five Finnish educators were presented. The research data were organized according to the five central themes that emerged from the data analysis process. Most notably, teachers in both Finland and New York State commented that when they were provided feedback on their teaching from a respected and trusted leader, the feedback was impactful and taken into careful consideration. Additionally, teachers in New York State and Finland were in agreement that the use of students' scores on high-stakes standardized assessments as a means to evaluate teacher effectiveness was an unfair, unhelpful practice that was based on erroneous assumptions. New York State teachers in the sample of this study do not wish to continue the practice, and Finnish educators in the study sample were in agreement that they do not wish to see the practice implemented in Finland.

While chapter four had a focus on the presentation of the research data and the five key emergent themes, this chapter will now review the data from a broader lens. Two additional, more comprehensive themes will be presented and discussed. The theoretical framework and relevant literature for this study will be revisited and connected to the research data. Finally, two of the research questions for this study, as well as the two supplemental research questions for this study will be addressed. The third research question on policy implications will be addressed later on in chapter six.

The Widening Scope of Expectations Regarding the Role of the Teacher

One broader theme that emerged outside of the bounds of the specific research questions for this study was the idea that the expectations regarding the role of the teacher are evolving. Teachers in New York State and Finland alike expressed concern over the widening scope of the role of a teacher, particularly regarding the degree to which teachers are expected to take ownership over their students' wellness. This discussion came up in the literature as well, where Webb et al. (2004) alluded to the role of the teacher as expanding to include a growing social work dimension.

Teachers know their responsibility to help students learn, and as explained in the literature and the research data teachers are naturally motivated by the moral rewards associated with the feeling of helping their students (Santoro, 2011). In teacher preparatory programs, such as in Educational Psychology class, teachers learn about Maslow's Hierarchy of Needs, or the theory that students' basic needs must be met before they are able to learn (McLeod, 2007). Having been educated in this theory, when students arrive at teachers' classes hungry, upset, worried, or with any type of unmet physiological or safety need, teachers face a pressing dilemma. They are faced with trying to meet those needs or attending to teaching and learning. Teachers sometimes take on responsibilities outside of their traditional roles, such as counseling students or keeping snacks in their classrooms for hungry students just so their students can learn. Teachers understand the importance of trying to efficiently meet the unmet need so that students can move "up Maslow's pyramid" to a place where they can focus on learning. This teacher dilemma of attending to either teaching, or the physiological or safety needs of students is especially exacerbated in New York State with the added judgment placed on teachers when their students do not achieve satisfactorily on high-stakes standardized assessments, which affects teachers' personal APPR evaluation scores.

Many of the sampled teachers in both New York State and Finland, at different points in their interviews, expressed concerns over the widening scope of expectations of the role of the teacher. Specifically, teachers are often on the front lines of mental health, family, and poverty issues with students. There is agreement among teachers in both New York State and Finland that these concerns are on the rise, and teachers feel that they no longer just "teach," but that they are expected to solve societal problems and that the lines between teaching and social work are blurring. Teachers in my sample also brought up that this expanding scope and increased responsibility could be contributing to teacher ineffectiveness.

Even though Finland has more social programs built into its governance structure than New York State or the United States, Finnish teachers also expressed concern about the idea of "too much is too much." They described the impact that the increase in expectations is having on their energy and inner growth, such as having to meet diverse student needs and learn new methods of teaching being presented by the government. They also talked about the wide range in ability within their student populations in their classes, and how they have to make instructional choices about the content and delivery of instruction in response to this.

> We have so much. This is really hard to explain. We have… we have many things, and we have a pressure to develop all kinds of systems at our school… So too much is too much. We are now discussing, if we can limit or say, "No. Stop. We don't want that anymore." …. And that is what is stopping my inner growth sometimes. Yes. Yes. Yes. Because we are tired. Sometimes, really very tired. I have now only 17 hours [to teach per week]. So, I can handle, because I have also Thursdays free from school. Yeah, but otherwise, if I… if I was all week and days there at school and if I do whole days there, whoa. That's too much for me. [Nora, Finnish Teacher]

> I mean we are like a basic… we should give the basic knowledge for the students. Very much from outside different organizations, they want to come and present, how this should be done, how that should be done. Or the government is presenting new ideas, sometimes the city is presenting new ideas. It's fine, but it's quite overloading our work, because then there is a task. [Hans, Finnish Teacher]

> I have come to conclusion that less is more. Yeah. I have evaluated what is the most important things they should learn in my subject, what are they? And I concentrate on them. And then the

> rest. I leave it because there is no time. The groups are very, sometimes very polarized. There are very good students, and quite low level students…. So, it's hard to meet all of those needs sometimes. [Hans, Finnish Teacher]

Related to this, another factor that may be contributing to the widening scope of the role of the teacher is students' use of mobile phones. One Finnish teacher commented on this, and said that she and her colleagues are discussing this a lot.

> Oh, now I'm really thinking that group work, which I use nowadays a lot. So, I see how are they [students] are working and what are they doing. And also, we write to Google Classroom, we have that. So, I see all the time what they are writing. If I want to, I can go and watch what they are doing…. . But of course, we have in Finland problems sometimes with mobile phones. Sometimes they are allowed to use them in the classroom sometimes, and sometimes not. And sometimes we can have problems with that. We are talking about that quite a lot. [Nora, Finnish Teacher]

One Finnish teacher talks about what he had to learn, in order to avoid burnout in the teaching profession.

> But long days, especially in the beginning, I really took every student's life to heart and it would keep me up at night. And I've really had to learn from the beginning to let that [stuff] go. At the end of the day, it is not my responsibility to save that kid's life. It would be great, but not my job. And that's Finland, because there's a social system. There's a very clear path for a teacher to take to communicate with city and social workers to get things figured out. [Kevin, Finnish Teacher]

The previous quote also nicely illustrates a key difference between schools in Finland, and schools in New York State. Because Finland has several social programs in place to ensure that families have their basic needs met by the government, such as food and health care. Kevin alludes to this in his quote above when he says, "there is a very clear path for a teacher to take to communicate with the city and social workers to get things figured out." This excerpt is also an important piece of data, because it shows that even in Finland where there are government systems established to address unmet safety and physiological needs of students, teachers like Kevin still grapple with what the boundary is within the role of the teacher, and the idea of it "not being their job to save that kid's life."

One Finnish teacher nicely illustrated the wide scope of the role of the teacher, and the realities of what teachers face every day in the classroom in order to be effective.

> But we're talking about the difference between a teacher and an educator, and the concept of like, teachers having multiple roles and wearing different hats all the time. Sometimes you're a police officer, sometimes you're a dad, sometimes you're an educator. You have to know the difference. [Kevin, Finnish Teacher]

On the topic of ineffective teachers, one New York State teacher commented that teachers who used to be very effective are now very worn out.

> [Regarding ineffective teachers] I don't know. I would say it's, it's got to be a problem. I think what more you're going to see are teachers that are just worn out. It's too much on them. And a lot of them stopped. I mean, the last couple of years, we've had so many retirements. Really, incredible teachers, but I think you're wearing out the teachers that spend a lot of time with the kids, a lot of time working on things, that are at school until 5:00 at night because they're grading papers or whatever. Those are effective teachers that are worn out. And I think I think it's really hard. [Lana, NYS Teacher]

Another New York State teacher commented that teachers like to feel they are in control of their students' learning, and the fact that so many factors lie outside of the control of the teacher causes teachers internal stress and pressure.

> I think the whole idea of finding a balance between personal and professional lives. I think that's what a lot of teachers have issues with, because we'd like to be in control so much, and that we can't shut it down. [Miranda, NYS Teacher]

Another New York State teacher commented that supervisors will bring up concerns about struggling students, yet they often do not have any specific recommendations or suggestions as to how to remedy the situation. This may reflect that they are unsure of what to do as well.

> How do you modify your lesson, so then those students feel like they're not losing anything, that they're part of the general population? But most of them [administrators], I feel like most of the things, it's easy to say, you know, even supervisor will ask you, but they don't really provide you with any help or anything. You know, they just asked me to go figure this out. [Yael, NYS Teacher]

The same teacher went on to say that the reason for this could also be political. Sometimes, directives are provided from the top of the organization, and by the time they trickle down to the

supervisors who are directly overseeing the teachers, the theory behind the feedback does not easily translate to what teachers can realistically implement with their students, in practice.

> I mean, it's almost like that is where the political thing kicks in. I'm sure the supervisor also gets feedback. ... They get information from a supervisor. So, somebody gives you some information, and says relay this information to you know, the people who work under you, and then, he or she doesn't really have all the answers... [Yael, NYS Teacher]

The Impact of COVID-19 on the Teaching Profession

The COVID-19 pandemic was a global event that had far-reaching effects in all aspects of daily life for people across the world, and the field of education was no exception. While COVID-19 was not the focus of this particular research study, the consequences of dealing with COVID-19 came up repeatedly in this study. When collecting data for this study, the final question on both the teacher questionnaire (Appendix A) and supervisor questionnaire (Appendix B) asked participants what effect, if any, did COVID-19 have on any of what was already discussed during the interview.

One finding that four different educators in this study brought up was the idea of an increase in the mental health needs of students. There have been significant gaps in learning, increases in depression and anxiety, and a decrease in students' ability to following directions.

> We've seen dips, very significant dips of learning loss, particular key grade levels and ages, just because some of the skills weren't able to be taught with the same rigor, the same level of expectation, and that creates challenges of, how you recoup two lost years. So, I think it definitely has impacted the cohort of kids that are going through the system, as well as new educators that are coming in, and haven't really seen a new norm fully and completely of what a new school year would look like. [Brian, NYS Supervisor]

> We have hired quite a few counselors to kind of help the kids through a lot of their anxiety and depression issues going through COVID. So, there's a lot of changes. [Lana, NYS Teacher]

> It seems like it's… really difficult for them to focus, and to keep the focus, and it's difficult for them to follow directions. It's difficult for them to even, like, read what they're supposed to do. [Linda, Finnish Teacher]

> Yes. This is sad news. I think these news over the world, that we have really find out that the seventh-grade students now, they are extremely… never we have had so many students who need so much extra help. So, I have this case that in one class, there's 24 or 23 students, and 10 need really a lot of extra help. That's really much to one teacher. [Nora, Finnish Teacher]

Additionally, teachers and supervisors in New York State commented that one positive effect of COVID-19 on the teaching profession was a decreased emphasis on students' scores on state standardized assessments. Teachers appreciated having the increased time and ability to focus on other aspects of their profession, beyond the demands of APPR.

> So, I guess in a way, it's had a positive impact because there's even less of a focus on our whole APPR… I don't know if I want to say performance, but the way that things were anyway. So, maybe that's one of the only positives to come out of all this. [Miranda, NYS Teacher]

> COVID certainly impacted, because the state said… basically said, don't worry about those professional scores, but it also I think, in a positive way, may have gotten people to think beyond the Regents exams. [Stephen, NYS Supervisor]

> I had the time during COVID to read books, to do some research, to learn more about my students. and then, when I went back I was like oh, maybe I can do this. Maybe we should do this, maybe I can add these kinds of things. So, it wasn't all negative. [Yael, NYS Teacher]

Furthermore, a few educators commented that the disruption to daily life that COVID-19 presented had an effect on sparking reflection and accelerating change in schools that might not otherwise have occurred.

> So many students were already behind, academically and socially. And so, COVID impacted not only the students, but the teachers, and the parents. So, it's almost like everybody at the same time. The teachers had to learn different ways to teach. Students had to learn different ways to learn. The impact was… I felt like the impact was enormous. [Yael, NYS Teacher]

> But what I think it has positively impacted is that teachers have had to reflect on their work differently. Because during COVID we're doing everything remotely, you have to think about your

> job in a new way. Many subjects they really have to think about what's actually important. [Lily, Finnish Teacher]

Overall, the effects of COVID-19 on the teaching profession have been broad and far-reaching. The impact on student learning, student wellness, educator learning, educator wellness, and the teaching profession as a whole has been profound, with some of the reflections represented in the data. While educators shared their perceptions in this research process, the long-term effects are still yet to be determined, and remain a subject worthy of future study.

Teaching as Art and Craft

In chapter two, the conceptual framework for this research study was presented. This framework, illustrated below, depicts teacher evaluations and the different factors that influence teachers' professional motivation to grow and develop.

Figure 5.1: Conceptual Framework for My Study
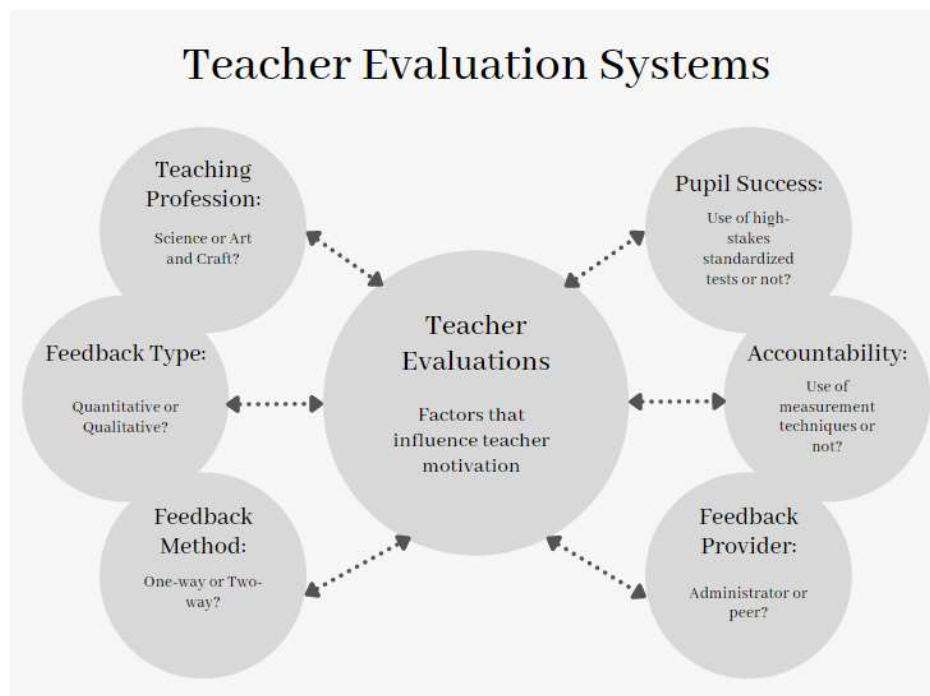
"I believe that many of the solutions being proposed to cure what people believe to be educational ills, solutions such as minimum competency testing, state mandated evaluation procedures, and other legislative panaceas, to be fundamentally misguided. They were born of suspicion and tend to motivate by the stick"
(Eisner, 1984, p. 12).

Previously in the second chapter, Eisner's (1984), *The Art and Craft of Teaching*, was discussed as the theoretical framework that informed the development of this research study. Eisner presented the theory of the teaching profession as being that of an art rather than a science. Eisner's work allows the reader to visualize classroom conditions and environment with which the teacher is presented, and describes teaching as "being swept up in the task of making something beautiful" (p. 12). Teachers make numerous impromptu decisions each minute, based on the information they are receiving and processing from their students and classroom conditions. Applying this to APPR and teacher feedback, Eisner's theory would argue that instead of seeking scientific solutions to educational problems, we should instead view schools as fluid, professional communities and focus policy efforts on creating conditions for teachers to continuously grow as professionals. Nurturing such conditions and creating opportunities for teachers to feel they have used their talents to positively impact their students' growth and development has been shown to be intrinsically motivating to teachers. This form of intrinsic motivation is much more powerful than extrinsic motivators for teachers, such as standardized evaluation practices like APPR (Eisner, 1984). This theory of action is more in line with the teacher-supervisor conversational feedback and professional growth practices that take place in Finland.

An elaboration of Eisner's (1984) theoretical framework that informed this research study was Meyer's (2016) *The Limits of Measurement: Misplaced precision, phronesis, and other Aristotelian cautions for the makers of PISA, APPR, etc.* Meyer (2016) offers the premise is that the classroom is a domain of practical knowledge, or phronesis, where quality is best appraised by

experienced practitioners. This is because practitioners possess the requisite context-sensitive judgment to provide feedback that is meaningful, appropriate, and free from extraneous interest. When quality is assessed by those who do not have the context-sensitive judgment that experienced practitioners possess, it begs the question of concerns over extraneous interest in education. Meyer (2016) calls for a restoration of an alternative conception of professional practice of education as a case of practical knowledge. The educational model in Finland more closely resembles this than in New York State.

One point to draw from the theoretical framework is that the person who delivers the feedback to teachers matters. It should be a school leader who is trusted and respected, who has a relationship with the teacher, who has a level of commitment to improving things, and who is in tune with the goings-on in the teacher's classroom and school building. As we saw in the data presented in chapter four, wise leaders' feedback is respected, impactful, and motivating. There is also a moral aspect of practical knowledge and that events must be co-constructed by an open and empathetic colleague. Applying this to the APPR teacher evaluation system, the use of rubrics and standardized test scores as a means to evaluate teachers is misplaced and unhelpful.

The other point to draw from the theoretical framework is that teaching is an art, not a science, and teachers are intrinsically motivated by the moral rewards of feeling like they are helping their students. They are not motivated by extrinsic rewards, or punishments, such as summer vacations or poor APPR scores. Santoro (2011) also presents the importance of teachers being able to access moral rewards in order to avoid burnout in the profession. These points from this literature, coupled with the research data, will be used to inform the subsequent discussion, and also formulate policy implications in chapter six.

Finland v. New York State

In response to my first research question on differences in teacher evaluations between Finland and New York State at the structural and cultural level, this study makes clear that at the structural level, New York State APPR teacher evaluation practices take the form of rubrics and involve extensive data analysis of standardized assessment results, classroom observations, and student growth in learning goals. Feedback is provided by the supervisor to the teacher in the post-observation conference in both written and verbal forms. The overall feeling is that teachers have to prove themselves to their supervisors and school communities that they are doing their jobs. In contrast, in Finland, there is no formalized teacher evaluation process. Instead, teachers are provided feedback by their supervisors and colleagues informally through two-way conversation, as the need and opportunity naturally arises in the profession. There is very little administrative oversight of teachers in Finland, and they are largely autonomous and trusted to do their jobs.

This study also makes clear the cultural differences in attitudes toward the teaching profession in Finland and New York State. Through the responses of the five educators in Finland and the six educators in New York State in this study, the reader can get a sense of the culture in each location, and that educators in Finland carry themselves differently than those in New York State. In general terms, educators in Finland operate within a culture of professional accountability where there is immense trust in teachers, and teachers are given a great deal of autonomy. In contrast, teachers in New York State operate within a culture that is driven my market accountability, including an expectation from the public on teachers to produce results and make demonstrable progress that can be quantified in real terms, such as that of standardized test scores that are suitable to the community. These cultural attitudes are reflected in the teacher evaluation practices and policies in Finland and New York State.

There are differences in governance and societal attitudes as well. In Finland, teaching is regarded by its citizens as one of the most prestigious occupations, and teacher education programs are highly competitive and available to only the top students through Finnish universities (Kelleher & Kase, 2012). Additionally, education is funded by the federal government, and not through a property tax structure as it is in New York State. Notably, even though education is funded by the government, and education legislation is prepared by the Ministry of Education and Culture, decisions about education governance are largely left to local school officials (*Finnish National Agency for Education*, n.d.). Sahlberg (2011) presents several lessons from Finland's success, including the creation of a respected profession in which teachers have a large degree of authority and autonomy, including responsibility for curriculum design and student assessment, and this engages them in the ongoing analysis and refinement of practice. New York State is highly regulated at that state level, particularly in curriculum and assessment, and by the incorporation of the practice of rating teachers based on their students' achievement, APPR evaluation policies reflect an underlying distrust in teachers. APPR rubrics in New York State also label teachers as either "ineffective, developing, effective, or highly effective." According to Firestone et al. (2014), teacher autonomy is one of the critical conditions in schools that maximizes intrinsic rewards, thus fostering teacher motivation, and this is more present in Finland than it is in New York State.

How do Teachers Grow?

In response to my second research question on how teachers in New York State and Finland view the contribution of teacher evaluations to their motivation and professional growth, it was surprising that neither New York State teachers nor Finnish teachers in my sample mentioned evaluations as contributing to their motivation nor their professional growth. Instead what emerged was the prominence placed on meaningful conversations with colleagues, as well as feedback that

was delivered by trusted and respected school leaders. This previously came up in the review of the literature when Arneson (2014) talked about trust between the teacher and his or her evaluator being necessary in order for feedback to be processed and accepted.

When asked about feedback that has been helpful or influential in their personal growth and development as professionals, Finnish educators in my sample talked about the care they have for one another as colleagues, and how the discussions have positively impacted them.

> Like, between the lines, feedback from students. Also, boss and from colleagues, it can be straight or straight paper, or not. But we have that kind of culture I feel that we are in a positive way. We care about what is happening and we are really sensitive to feel… sensitive to feel also to help each other… others at school as a colleague. So, we are talking all the time, how is it going? And I feel that we will have a really easy help if we really need. But I hope that everybody is enough open to ask for help. I hope that teachers are so couraged. [Nora, Finnish Teacher]

> But like I say it's been a mixture, there's nothing better than positive feedback. No, not just from your boss, but also from your colleagues and your peers. And it's nice to hear it. Guardians, and that gives one confidence then that… Yeah, I did this right. Oh, yeah, that was a good decision I made, and then from that you use the experience thereafter. You draw on that experience. So, in a way it is positive feedback. The self-reflection does come with time. I mean, one should never expect to be the finished article. [Ron, Finnish Supervisor]

> I don't know others, but it's very little I get feedback from my supervisors. Because my class is going fine…. But mostly from colleagues. From colleagues. After university, there is nobody who's looking at my teaching. Nobody's looking how I treat students. Not neither timing nor how do I do presentations. So, it's on your own then. But from my colleagues, also from courses. [Hans, Finnish Teacher]

> I think feedback from students. And then, like reflection, and just like the feeling you get working with the students. But also, so I believe that administrators last because I don't think that's been definitely that useful. But colleagues of course, the benefit of being slightly bigger school, of course, is that we have quite a few maths teachers and science teachers. So, we have a good group and we work together a lot. [Lily, Finnish teacher]

When asked the same question about feedback that has been helpful or influential in their personal growth and development as professionals, New York State educators in my sample once again talked about the value of exchanges with their colleagues. They also talked about the value of the conversation. This is of particular interest, considering the substantial APPR teacher

evaluation process that exists in New York State. Teachers and supervisors did not mention scores or rubrics is being influential at all. They repeatedly stressed the value of conversations, as well as other forms of influential development activities such as reading and attending professional conferences.

> Probably really my experiences with other teachers. I mean, for me anyway. You know, like when you're doing your own thing, experiences are great, but they're yours in isolation, you know, so I always liked that collaboration with other people, because there are some perspectives you just miss, you could be looking at the same thing and seeing it two completely different ways. So, you might not always agree, but I love the viewpoints you know, and sometimes my mind changes and sometimes it doesn't. [Leah, NYS Teacher]

> I don't know many teachers who are like, "I don't care what they [administrators] think." People do. Right? And being able to know what the issue is, is the only way you can at least try to address it. The conversations have been helpful, and they've made a difference. [Stephen, NYS Supervisor]

> I've really tried to seek professional development, going to the national conference, going to the state conferences, and also just really… I read a lot, and I think that's very important. [Miranda, NYS Teacher]

These findings also answer my second supplemental research question on the degree to which teacher evaluation practices contribute to professional growth in Finland versus New York State. Educators in my sample did not cite their personal growth as being stimulated by the scores they receive on evaluations. In fact, not a single teacher said their students' results on test scores, nor final evaluation designations were motivating in helping them improve. Instead, it is clear that positive feedback and meaningful conversations are of value to educators, and they feel that this contributes to their professional growth and development.

Additionally, while we know from the theoretical framework that teachers are naturally motivated by the feeling of helping their students improve and succeed, one key distinction to be made is that it is unclear as to what the specific catalysts are that motivate an educator to pursue professional growth activities. Though certain types of feedback and interactions have been

repeatedly cited in this study as being of value, it would be an interesting future study to take a closer look at what specific stimuli actually motivate a teacher to pursue activities to improve their craft.

On Teacher Satisfaction with Evaluation Practices

In response to my first supplemental research question on teacher satisfaction with evaluation practices in New York State versus Finland, the findings of the study demonstrate that there are differences in teacher satisfaction with evaluation practices in each location. Teachers in Finland do not have formal evaluation practices per se, however the Finnish teachers in my sample expressed in their interviews that they are largely content with the feedback delivery and content they receive from their supervisors. They often mentioned receiving only positive feedback, and in many instances I had to specifically ask them if they ever get any negative feedback because the examples they provided repeatedly described instances where they received positive feedback. When asked, the Finnish teachers in my sample had to really think about it, and the only examples of negative feedback were discussions between teachers and supervisors regarding the teacher's mental health and wellness, or regarding a specific parent complaint. Even then, the teacher was given the opportunity to participate in a discussion about the issue with their supervisor, and the interaction was perceived as fair.

Conversely, teachers in New York State in my sample expressed satisfaction in some areas, and dissatisfaction in other areas. Similar to the sampled teachers in Finland, the sampled teachers in New York State expressed satisfaction with their discussions with their supervisors when they felt like their supervisors knew them well, and respected them as teachers. Many expressed that these discussions were beneficial, and many also referenced positive feedback as being helpful.

However, New York State teachers and supervisors in my sample expressed dissatisfaction over the portion of the teacher evaluation process that uses students' standardized assessment results to evaluate teacher effectiveness. Teachers and supervisors perceive this process as unfair, inappropriate, and unhelpful in their growth and development as teachers. Finnish educators in my sample also expressed concerns with this practice in their interviews, and were united in the sentiment that they do not wish to see this practice come to Finland.

Summary

This study demonstrates the cultural differences in attitudes toward the teaching profession between the Finnish and New York State education systems, and how these differences present themselves in teacher evaluation practices. This qualitative study has thus far answered two primary and two supplemental research questions, and identified five emergent themes that can be used to inform future policy efforts in the area of teacher evaluations. The third primary research question on policy implications will be answered in chapter six.

In chapter six, universal themes beyond the scope of this study will be presented and discussed, as well as extensions of this topic. Four key policy implications will be outlined as a result of the research study findings and discussion.

CHAPTER 6: POLCY IMPLICATIONS AND CONCLUSIONS

Introduction

This research study provides a comparison of teacher evaluation systems in New York State and Finland at both the structural and cultural levels. Additionally, we learned that it is not a formal teacher evaluation process that contributes to teachers' professional growth per se, but rather the content and quality of the conversations in which teachers participate with their colleagues, students, and supervisors that has the greatest impact. These conversations may take place with supervisors as part of teacher evaluations, however it is the conversations occurring outside of APPR, or formal processes, with colleagues and students that teachers cited as being most impactful on their own professional growth and development.

As we saw in the research study data in chapter four and discussions in chapter five, the teacher evaluation systems between the two countries are different in both structure and culture. Finland does not have a formal structure for teacher evaluations, whereas New York State has state-approved rubrics and formulas. Additionally, there are cultural differences in the way teachers are perceived by the public in Finland versus New York State. In chapter six, the research study will be reviewed holistically and summarized. Policy implications will be presented, limitations of the study will be outlined, and recommendations for future research will be offered.

Summary of Main and Incidental Findings

As discussed in chapter four, there were five distinct themes that emerged from the research data collected from Finnish and New York State educators. The five major themes identified from the results of this study included:

1. Teachers appreciate feedback from supervisors whom they have a relationship with, and personally trust and respect.

2. Teachers and supervisors feel the use of high-stakes standardized test scores as a means to evaluate teacher effectiveness is an unfair and inappropriate practice.

3. Teachers cited many sources of helpful or influential feedback in their professional growth and development, including sources outside of a formal evaluation process.

4. There is a cultural difference in the way teachers and the teaching profession are perceived in New York State versus Finland.

5. The majority of teachers are very good, and the problem of the "ineffective teacher" is small.

In addition to the five emergent themes, two broader themes also came to light during the data analysis process:

1. The scope of expectations of the role of the teacher is widening.

2. The COVID-19 pandemic had far-reaching effects on the teaching profession, including student learning, student wellness, educator learning, and educator wellness.

The most prominent findings of the research data were in the area of feedback that is meaningful and impactful to teachers. We learned the importance and impact of feedback that is provided from wise and committed leaders whom teachers personally know, trust, and respect. We also reaffirmed the sentiment among teachers in New York State being the practice of using students' scores on high-stakes standardized tests as part of the evaluation process is unfair, unhelpful, and based on flawed theories of change about the teaching profession. This research study also affirms that Finnish teachers are not interested in this practice and would not like to see it come to Finland. Webb et al.'s (2004) research supports this notion as well, explaining that when

governments are overly prescriptive, this undermines public respect for teachers because it conveys a message that "teachers are people who need telling what to do" (p. 92).

Many forms of feedback are meaningful and impactful to teachers, including those that lie outside of formal evaluation practices such as APPR. In fact, teachers talked more about interactions with their colleagues and students, as well as their own professional reading and personal reflections as being impactful. No teachers cited APPR rubrics or students' standardized test scores as being impactful on their teaching.

Additionally, there is a cultural difference in the way teachers and the teaching profession are perceived in New York State versus Finland. In Finland, teachers have a lot of trust and autonomy, and there is no formal feedback or evaluation process other than natural conversation between the supervisor and teacher. In New York State, teachers are largely evaluated on their quality based on how their students perform on high-stakes standardized tests, and their teaching quality is assessed by their supervisors using a variety of methods, including numeric scores on rubrics.

Teachers and supervisors in both New York State and Finland were in agreement that the majority of teachers are good, and the problem of the "ineffective teacher" is small. This is important, because if the problem of the ineffective teacher is small and the vast majority of teachers are trustworthy and do good work with children, then a subject worthy of inquiry is whether it is wise to have a major policy reform designed to remedy a small problem.

Does Finland offer an Alternative to APPR?

Does Finland offer an alternative to APPR? This study suggests that it does. In Finland, a collaborative, communities of practice model is utilized that works well, and a version of that could be used in the United States. There is no direct equivalent to APPR in Finland; no observation

visits, no high-stakes standardized tests, no rubrics, no formal evaluation process. Instead, there are trusting relationships between collaborative colleagues.

It is important to acknowledge that anyone being observed by their supervisor is in a vulnerable position. They find themselves in a situation where they are opening themselves up to an outsider looking in on their classroom, which tends to cause the teacher to become defensive. To be able to receive feedback that is in the other person's good intentions is essential. In the next section, specific policy implications are outlined as a result of this study's comparison of evaluation systems in New York State versus Finland.

Policy Implications

In response to my third research question on policy implications, this study suggests that there are alternatives to APPR, such as the case of Finland, and that there are forms of feedback that are more motivating and impactful to teachers than others. Using these to inform teacher evaluation policy will improve the field at large. Broadly, there is a call for more authenticity in the APPR teacher evaluation policy in New York State and beyond, and a need for more meaningful connection between education professionals.

1. **Invest in Collaboration, Peer-to-Peer Learning, and Communities of Practice Between Professionals.** Because teachers in New York State and Finland in this study repeatedly cited the value of conversations with their colleagues, as well as feedback from trusted and respected leaders, more opportunities for two-way dialogue between professionals is the first policy recommendation. This could be in the form of facilitating more communities of practice, providing teachers with more time in the work day for peer discussions and/or peer observations, or creating more opportunities for colleague-

colleague or colleague-supervisor collaboration (Wenger & Snyder, 2000; Darling-Hammond et al., 2010). In addition, dialogue between supervisors and teachers should be frequent and ongoing as well. This helps the teacher feel supported with his or her students, and it helps the supervisor stay in touch with the needs of the students, and be in tune with the goings-on within the school building at large (Hammerness et al., 2017).

2. **Detach Standardized Assessment Data Analysis from the Teacher Evaluation Process**. While students' performance on standardized assessments can be informative in the realm of curriculum and instruction, there is widespread agreement in the field that the practice of using these scores to assess teacher quality is an unfair, unhelpful, and inappropriate practice. It does not help teachers improve, and it rests on the faulty assumption that teachers are not doing their best (Webb et al., 2004). The practice of evaluating teachers based on their students' scores creates an environment where teachers do not wish to have low-performing students on their class rosters, for fear that they will be judged as low-performing teachers (Feng et al., 2010). Analysis of standardized test results can be insightful within the realm of curriculum and instruction, and this can be its own process outside of APPR teacher evaluations. In general, teaching is an art and craft, not a science, and this type of feedback operates on the underlying assumption of distrust of teachers, and is not helpful nor motivating in teachers' professional growth and development (Eisner, 1984).

3. **Shift the feedback from written rubrics and scores, to frequent conversational feedback.** One lengthy written document containing numeric scores is less meaningful than the daily conversations that take place between school leaders and teachers. In general, if a policy benefits the majority of people it impacts, it is considered to be a good policy

(Birkland, 2020). In the research data collected in this study, we learned that there are only a few ineffective teachers, yet it is the concern over these few ineffective teachers that is driving the current APPR policy for the entire teaching population. If very few teachers are considered to be ineffective, we should instead construct a teacher evaluation policy that reflects that the majority of teachers are naturally motivated to grow and help their students, and benefit from increased autonomy. As part of this, abandoning the lengthy rubrics and documentation process will free up time for principals and administrators so that they can be more present in their buildings and engage in "management by walking around" (Hammerness et al., 2017). The amount of time spent documenting APPR for all teaching staff is vast. Documentation procedures for the few ineffective teachers can instead take place outside of the APPR process, and can better target the specific ineffective behaviors that are occurring. The time spent documenting and writing things down for good teachers could then be repurposed on other things that improve schools (Firestone, 2014).

4. **Increase On-site Control over Teacher Evaluation Systems.** Teacher quality is best assessed by school leaders who have the requisite context-sensitive judgment to provide meaningful feedback for continuous improvement of schools (Firestone et al., 2014). The current APPR policy in New York State requires the use of state-approved rubrics as a means to evaluate teachers, and this feels rigid at times, and does not allow for the level of flexibility that is needed to effectively implement the policy when facing highly specific circumstances that are present in classroom or school community. Finland's culture of trust in teachers provides the necessary flexibility from a governance standpoint, and allows administrators to participate in meaningful conversations as a means to discuss concerns in ways that support the goals of the district or community.

On "Ineffective" Teachers

The research data in this study collected from educators in Finland and New York State alike revealed that both teachers and supervisors feel there are very few ineffective teachers. Furthermore, when there are ineffective practices occurring in the classroom, the research data revealed that teachers found it most motivating and impactful when feedback was provided by a leader whom the teacher trusted and respected. Additionally, teachers prefer informal conversation about the concerns at hand, over written feedback given to them in a formal evaluation document.

This research data has implications for future policy efforts in the area of teacher evaluation in a few key ways. First, the fact that there is a common understanding among professionals in the field that there are very few ineffective teachers further supports the need to pivot from a theory of change that presumes educators are not doing their best, to a theory of change that presumes the vast majority of teachers are doing all they can to support their students' success. Second, if the purpose of teacher evaluations is to help teachers grow and develop professionally, a feedback model should be adopted that has been shown to be motivating and meaningful to teachers. Based on the data in this research study, such a feedback model would include feedback delivery from either a colleague or a respected and trusted leader, and feedback would take place in the form of conversation, rather than a formal written document. It is notable that when teachers in this study were asked about feedback that was helpful or impactful to them, none of the interviewed teachers cited APPR forms or formal written observation processes as being helpful.

It is worth noting that if serious concerns arise with a teacher, there are written documentation processes that exist outside of APPR and within the human resources area of school district oversight that can be followed. This paper is not calling for the elimination of staff member discipline, it is instead making the case that there are alternatives to the APPR teacher evaluation

system, and that the extensive written rubric documentation that is currently taking place for all teachers in New York State might not be necessary, when there are only a small number of ineffective teachers. In addition, other processes exist outside of APPR to address ineffective behaviors or teaching practices that may be occurring in the classroom.

Theory of Change Behind Education Reform in New York State

As part of summarizing the findings of this study and understanding its implications for future policy reform efforts, it is important to reflect on the historical and political context of the current APPR policy that exists in New York State. In the past 25 years, the three leading approaches to education reform in the state have been standardized test-based accountability, state-mandated teacher evaluation systems, and increased school choice, such as through charter schools. Each of these school improvement methods is based on an underlying theory of change that schools and school leaders are not doing their best, and require extrinsic motivation, such as market accountability pressure, in order to improve.

The findings in this research study speak to flaws in this theory of change. Eleven educators were interviewed in this research study, five being from Finland and six being from New York State. When asked about what motivates them to grow and develop, all eleven shared various sources of motivation, such as feedback from colleagues, students, supervisors, or their own professional reading and personal reflection. None of the eleven educators cited increased accountability pressure as being motivating nor helpful in their growth and development as educators. In fact, when the Finnish educators were asked how they would feel if such accountability approaches were introduced in Finland, all five admonished such practices and used words such as "devastated" to describe their feelings on the matter. Another said that doing so,

"would go against everything we believe in." A reminder that Finland was chosen for comparison in this study because of its highly literate society and globally-recognized and well-developed education system (Sahlberg, 2011; Schwab & Zahidi, 2020). Finland operates on a professional accountability model, where there is a tremendous amount of trust in teachers and they are given a great deal of professional autonomy (Williams & Engle, 2013).

In the theoretical framework for this study, Eisner (1984), has shown that educators are intrinsically motivated by their work, and in particular the "moral rewards" associated with the feeling of making a difference for their students. As Eisner (1984) put it, nurturing such conditions and creating opportunities for teachers to feel they have used their talents to positively impact their students' growth and development has been shown to be intrinsically motivating to teachers and exceeds whatever motivation it is that sabbaticals or vacations can provide (Eisner, 1984). Meyer's (2016) work adds that the classroom is a domain of practical knowledge, or phronesis, which is tacit and resists scientific codification. Thus, educational quality is best assessed by practitioners who have the necessary context-sensitive judgment (Meyer, 2016).

As a result of the theoretical framework, research data, and analysis presented in this research study, this paper calls for a shift in the theory of change that has been being used to drive reforms in education in New York State. Instead of using a theory of change that schools and educators are not doing their best, we should instead embrace a theory of change that educators are naturally motivated by the feeling of helping their students succeed. Accordingly, the purpose of teacher evaluations in New York State should shift from one of measurement to one of teacher development.

How Teachers Change with Experience

One interesting finding to come out of this study was hearing the reflections of teachers in how they have changed as educators since beginning their careers. Almost all interviewees, in New York State and Finland alike, commented in some sense on how they were more "rigid" in the beginning of their teaching careers, and focused primarily on the curriculum. Yet, as they progressed and gained more years of experience in working with students, they developed more empathy for students and families, and have broadened their understanding of different circumstances their students may be facing, and used this to inform their instructional approaches.

> I've got, I think more empathy. Far more empathy with kids and parents, and understanding, I think that comes with being a parent and just seeing. For the kids especially these last several years, the struggles and the challenges. So, I think I am more empathetic and understanding towards kids and towards teachers as well. [Stephen, NYS Supervisor]

> Well, I've learned along the way, you have got to let go, and kind of go with the flow. Back in the day, it was a little bit harder for me [with respect to the curriculum]… So, I think that's changed a lot in 20 years for me. [Leah, NYS Teacher]

> I think I've become more interested in the students, holistically. If you'd asked me when I started working…. So, although I feel passionate about teaching them maths, I'm also kind of always aware that there are so many things that I think are much more important to their lives than what I'm actually doing with them. [Lily, Finnish Teacher]

This is an important finding, and is placed purposefully in this dissertation at this point. If experienced educators found that as they spend more years working with students, they continue to progress toward a more holistic approach to instruction, and their empathy for students and families only grows as the years go on… what does this mean for the makers of education policy? Seemingly the theory of change lying behind reform efforts in New York State in the last 25 years has been absent of an acknowledgement of the holistic and societal factors affecting students, teachers, and the field of education at large. Meyer's (2016) assertion of the classroom being a domain of phronesis comes to the forefront in this context, in that unless policy-makers have many

years of teaching experience themselves, or policies are written specifically by those who do, these individuals will lack the required context-sensitive judgment. Finland's culture of trust in teachers provides the necessary flexibility from a governance standpoint in that it is difficult to have a specific state policy in the realm of teacher evaluation, due the highly specific set of circumstances present in each classroom and school community. If any such state-level or federal-level policy on teacher evaluations existed, it would have to be broad enough to allow for the necessary flexibility and unique circumstances of each educational setting to achieve its intended purpose. This is why this paper is suggesting that a return to on-site control over teacher evaluation policy would be more appropriate.

Vision for Teacher Evaluations in New York State

Based on the policy implications of this study, the following is a vision of what a reformed evaluation system could potentially look like. Based on the research data, the purpose of teacher evaluations should pivot from one of measurement of teachers, to one of development of teachers. As part of this, practices that are motivating to teachers should be built into the teacher evaluation process, such as collaboration between colleagues, frequent two-way dialogue between teachers and school leaders, and meaningful professional development for teachers.

Wenger & Snyder's (2000) concept of communities of practice is gaining momentum in the field, and is worthy of consideration when developing future policy for teacher evaluation. Communities of practice are groups of people informally bound together by shared expertise and passion for a joint enterprise, and have been shown to be intrinsically motivating for teachers to improve. An example of a community of practice in the field of education would be a team of Algebra teachers who enjoy working together and collaborating on curriculum and instruction. Communities of practice are motivating, naturally sustainable, and have been shown to improve

organizational performance. They are also free from over-regulation, which supports the recommended theory of change that presumes educators are naturally motivated to help their students succeed (Wenger & Snyder, 2000).

In developing policy that supports teacher development as the main purpose of the teacher evaluation process, it should be considered how challenging it must be for teachers to become better teachers, without the ability to see other teachers teach. Peer learning and observation can be powerful tools to support growth, particularly for newer teachers who are still developing their repertoire of skills in their classrooms. Frequent, informal observations from trusted and respected supervisors are also beneficial in supporting teachers' growth. Additionally, meaningful professional development offerings that interest the teacher should be available and encouraged as part of supporting teachers' growth and development (Smylie, 2014).

From a policy standpoint, in order to support teacher's growth and development, teachers need time to collaborate, observe their peers, and participate in professional development. Thus, part of the implementation plan for this policy would be to allocate the necessary funding to hire additional teachers so that more time can be freed up in teachers' days to participate in professional growth activities, as part of advancing the field. School master schedules should be reconfigured to allow for more release time and professional growth activities for teachers, and in order for that to happen, more teachers need to be hired to teach and supervise the students during those repurposed instructional minutes.

It is not only about remedying the policy and allocating funds. There is also a significant teacher labor shortage in the field of education right now in New York State. Teaching is not seen as the prestigious, highly-regarded profession in New York State that it is in Finland. However, it would be expected that if steps are taken to remedy some of these policies that are negatively

impacting teachers, such as APPR, the use of high-stakes standardized test score accountability measures, and the lack of time for professional growth opportunities in the teacher work day, that this will send a message to those who are considering teaching as a profession that the state is heading in a direction that shows respect for teachers and is committed to improving the field of education. Taking these steps can only make teaching more attractive as a profession, and this would be important in generating enough interested teachers to staff the new positions and support these policy changes as part of committing to a shift in the investment in teacher growth and development in New York State.

Two Key "Tensions" at Play in Teacher Evaluation Policy

In order to best understand the full scope of the policy implications of this research study, there are two key tensions that need to be discussed. The first is a tension between teacher evaluations whose purpose is measurement, and teacher evaluations whose purpose is development. As mentioned in chapter two, Marzano (2012) explains that an evaluation system that fosters teacher learning will differ from one whose aim is to measure teacher competence. APPR was designed with the intent to measure teacher competence, but the theoretical frameworks for this study, along with ample other research in the field have demonstrated many of the practices being used in APPR to be flawed, such as the use of students' high-stakes standardized test scores to assess teacher quality. Future policy efforts in the area of teacher evaluation in New York State should pivot in the direction of having the purpose of evaluation be teacher development.

The second key tension that needs to be addressed as part of understanding the policy implications of this research study is that what is the best policy for an individual teacher, is not necessarily the best policy for the teaching profession at large. School leaders have long grappled

with this aspect of education policy-setting, as it can be challenging when there is a faction of people a policy is not serving well, no matter how small. In general, if a policy benefits the majority of the people it serves, it is considered to be a good policy (Birkland, 2020). This tension is not new, and in fact a few Finnish educators commented on this concept during their interviews.

> And that's also I think, a huge burden to know that whatever I do, I'm not going to reach all my students, because I can't teach them each individually. [Lily, Finnish Teacher]

> And then always you think that… which level I teach…. but I can't only teach the ones who can't intake it. I think also those who can. Because otherwise we are making hidden curriculum in classes. Yes, we try to follow that all these non-academic or students with difficulties to learn, that we try to fill they needs, not those who could go further on. [Hans, Finnish Teacher]

> I'm thinking that even my methods that I use in this school with regards to administration may not work in another school. Because the audience is different, the environment, the classroom dynamic, I mean, it's all completely different. [Ron, Finnish Supervisor]

Applying this to teacher evaluation policy implications, the current policy rests on the theory of change that teachers are not doing their best, and thus we need an instrument to identify weaknesses and address them. However, the theoretical frameworks used in this study challenge that theory of change, and we know that in fact teachers are naturally motivated by the feeling they get when they help their students succeed. Additionally, there are very few ineffective teachers, and to inform the entire APPR evaluation process for the state on the theory of change targeted at such a small subset of the teaching population is misguided. Ineffective teachers and ineffective teaching practices certainly need to be addressed as previously mentioned, however the steps to do this can take place outside of the teacher evaluation process.

Limitations of the Study

While studying the Finnish approach to teacher evaluation and its practices that foster intrinsic motivation in teachers is a worthwhile endeavor and could be potentially insightful in

informing policy efforts, there are many reasons why precisely copying the Finnish system in New York State would not be possible. As previously mentioned in the introductory chapter, Finland and New York State (in the United States) are distinct in many ways. Some of these key differences include their governance, geography, demographics, and culture. For example, Leonardatos (2015) found that principals in New York State have less autonomy than Finnish principals. Thus, even if policymakers New York State were to reform APPR to "match" the Finnish approach to teacher evaluation, there are many factors at work that could potentially undermine its effectiveness in New York State.

Additionally, in this research study, I wish I could interview even more people. If it were feasible in qualitative research to interview hundreds of people, I would be curious to see if the emergent themes in this study continued to hold true. It is often said that the shortcomings of qualitative research can be answered by quantitative research, and vice versa. It would be great to take this study, and design a quantitative companion study where data could be collected on similar topics from the masses, and analyzed using quantitative methods.

Recommendations for Future Research

Looking deeply into the topic of teacher evaluations has sparked my curiosity, and lead to a few topics that are potentially worthy of further inquiry in the field. One of the things we do not understand well enough is what motivates teachers to grow as practitioners. Is it feedback from supervisors? Is it collaboration with peers and/or participating in communities of practice? Is it intrinsic interest in helping students and/or the subject matter? Is it having the freedom to experiment in the classroom, and participate in the trial and error process often associated with delivering lessons to students? Looking more closely at this area would be beneficial to the field, and further inform efforts to develop systems that promote the growth and development of

teachers.

Another idea worthy of further study would be to look at alternatives ways to fund school systems, as opposed to the property tax system that is utilized in New York State. I believe funding schools using a system partially driven by the ability of local taxpayers to contribute has several negative consequences for school districts. Additionally, it is possible this funding system may intentionally or unintentionally reinforce theories of education reform that presume negative intent of teachers at times.

Another idea worthy of future study would be to look at the ways COVID-19 has changed education, and how this can inform policy efforts in the area of curriculum and high-stakes standardized test administrations in New York State both short-term and long-term.

Concluding Remarks and Reflection

The United States is culturally and geographically distinct from Finland, but can still learn from Finland's experience with educational change. Interviewing educators in Finland and New York State, and comparing their teacher evaluation systems and methods of feedback has been insightful. It has been interesting to hear educators share their thoughts on what has been helpful to them in their growth and development. This research study will hopefully inform future policy efforts in New York State, and elsewhere in the United States with similar evaluation policies.

The next recommended steps to improve teacher evaluation systems in New York State and beyond would be increased collaboration between professionals, a detachment of the practice of using state standardized assessment scores to evaluate teacher quality, a shift from a written feedback model to a conversational feedback approach, and overall increased on-site control over teacher evaluation systems.

Part of being a school leader is envisioning a better world for our children. It is also then taking actionable steps so that one day the vision will become a reality. It is my hope that by participating in the research community as a practitioner, and using research to inform decision-making, we can improve conditions in schools to transform our world so that our children will have a better future.

REFERENCES

Aldeman, C., & Chuong, C. (2014). Teacher Evaluations in an Era of Rapid Change: From "Unsatisfactory" to "Needs Improvement." In Bellwether Education Partners. Bellwether Education Partners.

Arneson, S. M. (2014). *Building trust in teacher evaluations: It's not what you say; it's how you say it*. Thousand Oaks, CA: Corwin Press.

Axe FX III | Page 80 | The Gear Page. (n.d.). Retrieved March 19, 2022, from https://www.thegearpage.net/board/index.php?threads/axe-fx-iii.1907452/page-80

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., and Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers.* Issue brief #278. Washington, DC: Economic Policy Institute. http://epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf.

Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education, 56*(3), 205-213.

Birkland, T. A. (2020). *An introduction to the policy process: Theories, concepts, and models of public policy making*. New York: Routledge.

Bradley-Levine, J., Romano, G., & Reichart, M. (2017). Teacher Leaders' Influence on Teachers' Perceptions of the Teacher Evaluation Process. *International Studies in Educational Administration (Commonwealth Council for Educational Administration & Management (CCEAM))*, *45*(3), 65-84.

Brinkmann, S., & Kvale, S. (2018). *Doing interviews*. Thousand Oaks, CA: SAGE Publications Ltd.

Bryk, A., Harding, H., & Greenberg, S. (2012). Contextual Influences on Inquiries into Effective

Teaching and Their Implications for Improving Student Learning. *Harvard Educational Review*, 82(1), 83–106.

Bryk, A. S., & Schneider, B. (2003). Trust in schools: A core resource for school reform. *Educational Leadership*, *60*(6), 40-45.

Callahan, R. E. (1962). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. Chicago: University of Chicago Press.

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*(3), 409-432.

Chaltain, S. 2012. "My View: When Did Teacher Bashing Become the New National Pastime?" CNN Schools of Thought, March 6. http://schoolsofthought.blogs.cnn.com/2012/03/06/my-view-whendid-teacher-bashing-become-the-new-national-pastime/.

Creswell, J. W. (2013). Qualitative inquiry & research design: Choosing among five approaches (3rd ed.). Los Angeles, CA: SAGE Publications.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry & Research Design: Choosing Among Five Approaches* (4th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Curtis, R. (2012). *Building it together: The design and implementation of Hillsborough County public schools' teacher evaluation system.* Washington, DC: Aspen Institute.

Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, *68*(4), 35-39.

Darling-Hammond, L., Armrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012, March). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15.

Darling-Hammond, L., Wei, R. C., & Andree, A. (2010). How high-achieving countries develop great teachers. *Stanford Center for Opportunity in Policy Education - Research Brief, August 2010*, 1-8.

Eisner, E. W. (1984). The art and craft of teaching. In Reinhartz, J. (Ed.), *Perspectives on effective teaching and the cooperative classroom: Analysis and action series* (pp. 19-31). Washington, DC: National Education Association.

Feng, L., Figlio, D. N., & Sass, T. (2010). *School accountability and teacher mobility* (No. w16070). National Bureau of Economic Research.

Finland-United States Relations - Wikipedia. (n.d.). Retrieved March 19, 2022, from https://en.wikipedia.org/wiki/Finland%E2%80%93United_States_relations.

Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal*, *44*(3), 594–629.

Finnish National Agency for Education. (n.d.). Retrieved January 15, 2023, from https://www.oph.fi/en

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, *43*(2), 100-107.

Firestone, W. A., Blitz, C. L., Gitomer, D. H., Gradinarova-Kirova, D., Shcherbakov, A., & Nordin, T. L. (2013). Year 1 report: New Jersey teacher evaluation pilot program. New Brunswick, NJ: Rutgers Graduate School of Education.

Firestone, W. A., Nordin, T. L., Shcherbakov, A., Kirova, D., & Blitz, C. L. (2014). *New Jersey's pilot teacher evaluation program: Year 2 final*. New Brunswick, NJ: Center for Effective School Practices.

Forman, K., & Markson, C. (2015). Is" Effective" the New" Ineffective"? A Crisis with the New York State Teacher Evaluation System. *Journal for Leadership and Instruction*, *14*(2), 5-11.

Gibbs, G. R. (2007). *Analyzing qualitative data*. In U. Flick (Ed.), The Sage qualitative research

kit. London: Sage.

Gigante, N., & Firestone, W. (2007). Administrative support and teacher leadership in schools implementing reform. *Journal of Educational Administration*, 46(3), 302-331.

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., Whitehurst, G. J., & Croft, M. (2011). Passing muster: Evaluating teacher evaluation systems. Washington: Brooking Institution, 1-36.

Haapala, H. (2017). Finnish schools on the move: Students' physical activity and school-related social factors. *Liikunnan ja kansanterveyden julkaisuja*, (336).

Hammerness, K., Ahtiainen, R., & Sahlberg, P. (2017). *Empowered educators in Finland: How high-performing systems shape teaching quality*. San Francisco, CA: Jossey-Bass.

Hansén, S. E., Forsman, L., Aspfors, J., & Bendtsen, M. (2012). Visions for teacher education– experiences from Finland. *Acta Didactica Norge*, *6*(1), (Art. 9, 17 pages).

Harris, L. R., & Brown, G. T. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 365-381.

Hess, F. M., & Willey, P. (2018). The limits of expertise. In Green, J. P., & McShane, M. Q. (Eds.), *Failure up close: What happens, why it happens, and what we can learn from it* (pp. 1-15). Lanham, MD: Rowman & Littlefield.

Hutchings, P., Borin, P., Keesing-Styles, L., Martin, L., Michael, R., Scharff, L., ... & Ismail, A. (2013). The Scholarship of Teaching and Learning in an age of accountability: Building bridges. *Teaching and Learning Inquiry*, *1*(2), 35-47.

Jennings, J. L., & Lauen, D. L. (2016). Accountability, inequality, and achievement: The effects of the No Child Left Behind act on multiple measures of student learning. *RSF: The Russell*

Sage Foundation Journal of the Social Sciences, *2*(5), 220-241.

Kelleher, J. & Kase, S. (2012). Teacher evaluation systems and multi-cultural commonalities and challenges: England, Mexico, Portugal, Finland, Ireland, and Turkey. American Educational Research Association, 2012 Annual Meeting.

Kamenetz, A. (2022, February 1). *More than half of teachers are looking for the exits, a poll says*. NPR. Retrieved February 13, 2022, from

https://www.npr.org/2022/02/01/1076943883/teachers-quitting-burnout.

Kansanen, P. (2003). Teacher education in Finland: Current models and new developments. *Institutional approaches to teacher education within higher education in Europe: Current models and new developments*, 85-108.

Kelly, K., Merry, J., & Gonzalez, M. (2018). Trust, Collaboration and Well-Being: Lessons Learned from Finland. *SRATE Journal*, *27*(2), 34-39.

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, *52*(5), 711-753.

Labaree, D. F. (2000). On the nature of teaching and teacher education. *Journal of Teacher Education, 51*(3), 228-233.

Leithwood, K., & Seashore-Louis, K. (2011). *Linking leadership to student learning*. San Francisco, CA: John Wiley & Sons.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. Minneapolis, MN: Center for Applied Research and Educational Improvement.

Leonardatos, H., & Zahedi, K. (2014). Accountability and "racing to the top" in New York state:

A report from the front lines. *Teachers College Record*, *116*(9), 1-23.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, *70(3)*, 14-19.

McLeod, S. (2007). Maslow's hierarchy of needs. *Simply psychology*, *1*(1-18).

Meyer, H. D. (2017). The limits of measurement: misplaced precision, phronesis, and other Aristotelian cautions for the makers of PISA, APPR, etc. *Comparative Education*, *53*(1), 17-34.

Meyer, P. (2011). Assessing New York's commissioner of education: with Steiner's sudden resignation, will the state continue its Race to the Top? *Education Next*, *11*(3), 32-40.

Milanowski, A., & Kimball, S. M. (2003). The framework-based teacher performance assessment systems in Cincinnati and Washoe (No. TC-03-07). Madison, WI: Consortium for Policy Research in Education.

Ministry of Education. (1999). *Education in Finland*. Helsinki: Ministry of Education.

Mintz, J. A., & Kelly, A. M. (2021). Science Teacher Motivation and Evaluation Policy in a High-Stakes Testing State. *Educational Policy*, *35*(1), 3-40.

New York State Education Department. (n.d.). *Rubrics*. New York State Education Department. Retrieved March 13, 2022, from http://www.nysed.gov/educator-quality/approved-teacher-and-principal-practice-rubrics.

New York State Legislature. (2010). Article 61. Teachers and Supervisory and Administrative Staff. Section 3012-c Annual professional performance review of classroom teachers and building principles. Retrieved March 13, 2022 from: https://newyork.public.law/laws/n.y._education_law_section_3012-c

Ni, Y., Yan, R., & Pounder, D. (2018). Collective leadership: Principals' decision influence and the supportive or inhibiting decision influence of other stakeholders. *Educational Administration Quarterly*, *54*(2), 216-248.

Patton, M. Q. (1990). Qualitative evaluation and research methods. Newbury Park, CA: Sage Publications.

Ravitch, D. (2021, November 30). *New Study finds that tougher teacher evaluations showed no gains in student achievement*. Diane Ravitch's blog. Retrieved December 1, 2021, from https://dianeravitch.net/2021/11/30/new-study-finds-that-tougher-teacher-evaluations-showed-no-gains-in-student-achievement/.

Sahlberg, P. (2011). Developing effective teachers and school leaders: The case of Finland. In Darling-Hammond, L., & Rothman, R. (Eds.), *Teacher and leader effectiveness in high performing education systems* (pp. 13-21). Washington, DC: Alliance for Excellent Education and Stanford, CA: Stanford Center for Opportunity Policy in Education.

Sahlberg, P. (2021). *Finnish lessons 3.0: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press.

Sahlberg, P., & Walker, T. D. (2021). *In teachers we trust: The Finnish way to world-class schools*. New York, NY: W.W. Norton & Company Ltd.

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education*, *118*(1), 1-23.

Schwab, K., & Zahidi, S. (2020). (rep.). *Global Competitiveness Report*. World Economic Forum. Retrieved March 13, 2022, from https://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2020.pdf.

Schwartz, B. (2011). Practical wisdom and organizations. *Research in Organizational Behavior,*

31, 3-23.

Seymour, C. M., & Garrison, M. J. (2016). What They Think about How They Are Evaluated:

Perspectives of New York State Physical Educators on Teacher Evaluation Policy.

International Journal of Education Policy and Leadership, *11*(6), n6.

Size of United States compared to Finland. MyLifeElsewhere. (n.d.). Retrieved March 19, 2022,

from https://www.mylifeelsewhere.com/country-size-comparison/united-

states/finland#:~:text=United%20States%20is%20about%2029,people%20live%20in%20

United%20States).

Snider, J. (2011). An interview with Henna Virkkunen, Finland's Minister of Education.

Hechinger Report.

Strauss, V. (2021, March 24). *Education secretary Cardona stands firm on standardized testing

mandate amid criticism*. The Washington Post. Retrieved March 12, 2022, from

https://www.washingtonpost.com/education/2021/03/24/cardona-stands-firm-on-

standardized-testing-mandate-amid-criticism/.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case

analysis of the connection between teacher effectiveness and student achievement. *Journal

of teacher Education*, *62*(4), 339-355.

Tarhan, H., Karaman, A., Lauri, K., & Aerila, J. A. (2019). Understanding teacher evaluation in

Finland: A professional development framework. *Australian Journal of Teacher Education

(Online)*, *44*(4), 33-50.

Taylor, F. W. (1967). *The principles of scientific management*. New York, NY: W.W.

Norton & Co.

Terrell, S. (2016). Writing a proposal for your dissertation: Guidelines and examples. New York,

NY: The Guilford Press.

Tirri, K. (1993). *Evaluating Teacher Effectiveness by Self-Assessment: A Cross-Cultural Study. Research Report 122*. Department of Teacher Education, Research Reports, PO Box 38 (Ratakatu 6A) 00014 University of Helsinki, Helsinki, Finland.

Tirri, K., & Puolimatka, T. (2000). Teacher authority in schools: A case study from Finland: JET. *Journal of Education for Teaching, 26*(2), 157-165.

Tjn. (2014, November 11). *View: APPR fails teachers*. The Journal News. Retrieved March 14, 2022, from https://www.lohud.com/story/opinion/contributors/2014/11/06/view-lower-hudson-council-school-superintendents-blasts-appr/18591403/.

Ulmer, J. B. (2016). Re-framing teacher evaluation discourse in the media: An analysis and narrative-based proposal. *Discourse: Studies in the Cultural Politics of Education*, *37*(1), 43-55.

Watt, D. (2007). On becoming a qualitative researcher: The value of reflexivity. *The Qualitative Report, 12*(1), 82-101.

Webb, R., Vulliamy, G., Hämäläinen, S., Sarja, A., Kimonen, E., & Nevalainen, R. (2004). A comparative analysis of primary teacher professionalism in England and Finland. *Comparative education*, *40*(1), 83-107.

Weber, M. (1946). *Bureaucracy*. In Gerth, H. H. & Mills, W. (Eds.), *From Max Weber: Essays in sociology* (pp. 240-244). New York, NY: Oxford University Press.

Williams, J. H., & Engel, L. C. (2013, January). How do other countries evaluate teachers? *Phi Delta Kappan*, *94*(4), 53–57.

Wenger, E. C., & Snyder, W. M. (2000, February). Communities of practice: The organizational frontier. *Harvard Business Review*, 139-145.

With nearly half of principals considering leaving, research urges attention to working conditions,

compensation, and supports. (2020, May 14). Retrieved April 27, 2021, from

https://www.nassp.org/news/with-nearly-half-of-principals-considering-leaving-research-

urges-attention-to-working-conditions-compensation-and-

supports/#:~:text=Concerns%20about%20principal%20turnover%20are,poverty%20school

s%20and%20rural%20communities.

Woo, H. (2019). Teacher Evaluation Policy Series | Professional Responsibility of Teachers:

Teacher Evaluation in Finland. *Forum of the American Journal of Education*. Retrieved

January 23, 2022, from https://www.ajeforum.com/professional-responsibility-of-teachers-

teacher-evaluation-in-finland-by-hansol-woo/.

Yli-Piipari, S., Kulmala, J. S., Jaakkola, T., Hakonen, H., Fish, J. C., & Tammelin, T. (2016).

Objectively measured school day physical activity among elementary students in the

United States and Finland. *Journal of Physical Activity and Health*, *13*(4), 440-446.

APPENDIX A

TEACHER INTERVIEW QUESTIONS

1. Tell me about your role in your school. What grade level(s) and/or classes do you teach? How many years of teaching experience do you have?
2. Do you feel that you have grown as a teacher in the years since you started? If yes, what has been most helpful in that growth? Your own self-reflection on what works in teaching? Exchanges with colleagues? Feedback from superiors?
3. What qualities do you think make up a great teacher?
4. How many supervisors do you have? What title(s) do they have?
5. Describe your experiences with receiving feedback on your teaching. Who provides you with feedback on your teaching? Where and when does this take place? How frequently?
6. What are some examples of feedback you have been provided? What was this experience like? What impact did this feedback have on you, if any? Was there any written documentation of the feedback provided?
   a. Any feedback on content knowledge? Instructional delivery? Student learning?
   b. Any feedback on relations with – colleagues, students, families, administration?
   c. Any feedback on attendance at work?
7. As you may know, in New York State, teachers have a formal system of being evaluated by their supervisors. Part of a teacher's final evaluation score rests on his or her students' performance on standardized tests. In Finland, there are no standardized tests, but if there were, how would you feel about using your students' scores on them as part of your evaluation?
8. Are there any other ways in which you are evaluated and/or provided with feedback?
9. To what extent does feedback contribute to professional growth?
10. How do you know how you are doing as a professional in your workplace, overall?
11. How satisfied are you with evaluation practices in your school?
12. To what degree do you feel teacher evaluation practices contribute to professional growth in your school?
13. Do you work with any ineffective teachers? What makes them ineffective? What course(s) of action exist, if any, to address concerns? Please describe them.

APPENDIX B

SUPERVISOR INTERVIEW QUESTIONS

1. Tell me about your role in your school. What grade level(s) and/or classes do you supervise? How many total years of administrative experience do you have?
2. Do you feel that you have grown as a teacher in the years since you started? If yes, what has been most helpful in that growth? Your own self-reflection on what works in teaching? Exchanges with colleagues? Feedback from superiors?
3. What qualities do you think make up a great teacher?
4. How many teachers do you supervise?
5. Describe your experiences with providing teachers with feedback on their teaching. Where and when does this take place? How frequently? Does anyone else provide teachers with feedback aside from you?
6. What are some examples of feedback you have provided to teachers? What was this experience like? What impact did this feedback have on the teacher, if any? Was there any written documentation of the feedback provided?
   a. Any feedback on content knowledge? Instructional delivery? Student learning?
   b. Any feedback on relations with – colleagues, students, families, administration?
   c. Any feedback on attendance at work?
7. As you may know, in New York State, teachers have a formal system of being evaluated by their supervisors. Part of a teacher's final evaluation score rests on his or her students' performance on standardized tests. In Finland, there are no standardized tests, but if there were, how would you feel about using teachers' students' scores on them as part of their evaluation?
8. Are there any other ways teachers are evaluated and/or provided with feedback? What is your role in this?
9. To what extent did the feedback contribute to the professional growth of teachers?
10. How do teachers know how they are doing as a professional in your building, overall?
11. How satisfied are you with evaluation practices in your school?
12. To what degree do you feel teacher evaluation practices contribute to professional growth in your school?
13. Do you supervise any ineffective teachers? What makes them ineffective? What course(s) of action exist, if any, to address concerns? Please describe them.