

Syracuse University

SURFACE

Dissertations - ALL

SURFACE

December 2020

Teachers' Perceptions of Evaluation Feedback Conferences: A Mixed Methods Study

Danielle R. Brain
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/etd>



Part of the [Education Commons](#)

Recommended Citation

Brain, Danielle R., "Teachers' Perceptions of Evaluation Feedback Conferences: A Mixed Methods Study" (2020). *Dissertations - ALL*. 1192.
<https://surface.syr.edu/etd/1192>

This Dissertation is brought to you for free and open access by the SURFACE at SURFACE. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Abstract

Teacher performance evaluations can serve two purposes: summative/accountability and formative/professional development. The current perception in the field is that performance evaluation systems predominantly focus on fulfilling a summative agenda over formative, which blurs the lines between the two purposes of evaluation (Popham, 2013). As a result, how evaluators and teachers react to evaluation ratings creates a disconnection between the summative and formative purposes and creates critical tensions between personnel being evaluated and evaluation systems. When this tension is felt, teachers and some evaluators feel that evaluation ratings cannot be used effectively for either purpose.

A way to lessen the tension would be for evaluators and teachers to focus on the part of the evaluation process within their control, the evaluation-feedback conferences. During feedback conferences, the evaluator and teacher discuss observations of the teacher's practice. This discussion, in theory, should be formative and summative for helping teachers at "improving instruction, ... assisting teachers to achieve their full potential, and improv[e] school culture and climate" (Willis & Ingle, 2015, p.71), and having teachers account for their own teaching decisions and the impact of their decisions on student learning (Peterson, 2004). The issue between which purposes feedback conferences serve raises questions about the impact of evaluation conferences over-all.

A body of research literature focuses on educational performance appraisal and observation process/protocols, but most of this literature focuses on how administrators should conduct classroom observations, approach evaluation conferences, and assign evaluative ratings. There is a paucity of studies that consider or explore teachers' experiences with *how* evaluators provide specific feedback from observations of practice, and how that feedback affects their

practice. There is a small body of literature that uses feedback theory to explain teachers' reactions to feedback, but that literature still shows a gap in understanding how teachers perceive the approaches evaluators use within the evaluation context when providing feedback on observations.

The purpose of this study is to describe teachers' experiences with evaluation feedback conferences and their perceptions of the impact those experiences have on their practice using a mixed-methods design. Analysis from qualitative data from interviews included in a Research Apprenticeship Project, quantitative data from an online survey on the dimensions of evaluation feedback conferences, and hybrid data (objective quantitative-subjective qualitative) from focus groups, all representing public school teachers who had an observation feedback conference with an evaluator, revealed teachers have complex, yet similar, perceptions of the evaluation conference experience. The data from this study has provided theoretical and practical considerations on how to conduct feedback conferences as part of an over-all evaluation system for teachers and evaluators that will have an impact teaching and learning, while also revealing the need for further research with a larger sample of teachers on the current directions evaluation feedback conferences across and between school organizations and districts in New York State.

**Teachers' Perceptions of Evaluation Feedback Conferences:
A Mixed Methods Study**

By

Danielle R. Brain

Bachelor of Arts, Utica College of Syracuse University, 1988

Master of Science of Teaching, State University at Binghamton, 1990

Certificate of Advanced Study: Teacher Leadership, Union Graduate College, 2010

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree
Of Doctor of Philosophy in Teaching and Curriculum

Syracuse University

December 2020

Copyright © Danielle R. Brain 2020
All Rights Reserved

Acknowledgements

...there is no such thing as teaching without research and research without teaching.⁵ One inhabits the body of the other. As I teach, I continue to search and re-search. I teach because I search, because I question, and because I submit myself to questioning. I research because I notice things, take cognizance of them. And in so doing, I intervene. And intervening, I educate and educate myself. I do research so as to know what I do not yet know and to communicate and proclaim what I discover (Freire, 1998, pg. 35).

The moment I read this by Paulo Freire in 2006, it became my mantra throughout the time I worked with Mohawk Valley Writing Project and then again through the last seven years working on this dissertation. I love research, conducting it and reading for it; research feeds my professional soul in ways I had not anticipated, as well as sustained me in ways that allowed me to persevere to get this dissertation done. As I approach the end, a special, heartfelt gratitude goes out to my dissertation committee, Dr. Leela George and Dr. Rob Pusch, for their patience, and my advisor, Dr. Joe Shedd, for his excitement and being ‘all-in’ on this project as we waded into methodology waters unknown to both of us.

One of the surprising outcomes of this dissertation project and program are the many colleagues and peers that have taken an interest in this endeavor. Those colleagues I can name, since they are not covered by confidentiality as research participants, have been mentors, supporters, and professional cheerleaders throughout this process. I need to acknowledge my greatest cheerleader, Katherine, who always seemed to sense when I needed a pep-talk. Without her support, I think I may have quit. Mentor Dr. MaryAnn Janda, my National Writing Project Director, who seemed to know this was my path and gently nudged me in this direction throughout our time with MVWP. A few colleagues at my ‘day job,’ always reminded me that such research pursuits are important: Dr. Krista Pembroke, since retired, forged the path and pulled me along with her on her PhD journey; Kathy Capozzella, mathematician and super-tutor for adult ‘math-phobics,’ reminded me that math is my friend and there is no crying allowed; my

school roomie, AnnMarie Farrell, and former department chairperson, Nadia Caleo, saw me through some anxious moments and always showed unwavering support when most needed.

The unnamed colleagues I must acknowledge are my research participants, for ALL my research projects the last seven years. I cannot name them all, but each one helped and supported any and all my research endeavors. I thank them for always saying, “of course, what do you need me to do?” and never once saying, “what are you doing that for?” All the participants opened themselves and their practice to me, without hesitation, which made all this possible.

Finally, I must acknowledge my best friend, spouse, and steadfast supporter, my husband Bill. Without his comment to me in July 2013, “I think it’s time you went for your doctorate” I would have never embarked on this professional journey. He may still ask what I am actually studying, how much longer it will take, and sometimes gets it mixed up when explaining to others, but he never once said to stop or quit. He has always had faith that I would get done.

⁵ One talks too much, with insistence, of the researcher teacher. In my opinion, research is not a quality in a teacher nor a way of teaching or acting that can be added to the one of simply teaching. To question, to search, and to research are parts of the nature of teaching practice. What is necessary is that, in their ongoing education, teachers consider themselves researchers because they are teachers (Freire, 1998, pg. 133).

Table of Contents

Abstract.....	i
Acknowledgements.....	v
Chapter 1	1
Introduction.....	1
Background on Evaluation Functions and Purposes.....	6
Problem Statement.....	11
Research Questions and Rationale.....	13
Theoretical Framework.....	15
Conceptual Framework.....	17
Methodology.....	20
Delimitation.....	20
Definition of Terms.....	21
Summary.....	24
Chapter 2	25
Literature Review.....	25
Literature on Theoretical Frameworks.....	25
Practical Literature: Evaluation Feedback Conferences.....	28
<i>Teacher Evaluation Systems' Function, Design and Structure</i>	30
<i>Dimensions of Evaluation Feedback Conferences</i>	36
<i>Feedback Impact on Teachers' Practice</i>	46
Summary.....	49
Chapter 3	51
Methods.....	51
Overview of Research Design, Instrumentation, Implementation, Data Collection.....	51
Research Focus and Questions.....	51
Research Study Design.....	54
Qualitative Paradigm: Naturalistic Inquiry—Semi-Structured Interviews.....	59
<i>Limitations</i>	64
Quantitative Paradigm: Survey Research.....	66

<i>Limitations</i>	83
Mixed Paradigm: Q-Methodology—Focus Group Q-Sorts.....	85
<i>Limitations</i>	93
Mixed Methods Validity.....	95
Personal Subjectivity Statement.....	100
Summary.....	104
Chapter 4	106
Research Findings and Analysis.....	106
Semi-Structured Interviews.....	106
Evaluation Feedback Conference Survey.....	116
Focus Groups: Q-Sort Data.....	147
Summary.....	166
Chapter 5	170
Discussion and Conclusions.....	170
Research Findings	171
<i>Narrative Data—Semi-Structured Interviews</i>	172
<i>Survey Data</i>	175
<i>Q-Sort Data</i>	179
<i>Center of Tension: The AFFECTIVE Connection</i>	182
Implications for Theory and Practice.....	185
Affective Evaluation Conference Theory.....	186
Praxis: Shift Towards Affective Feedback Conferences.....	188
<i>Teacher Agency</i>	189
<i>Formative-Summative Balance</i>	191
<i>Trust</i>	193
From Limitations to Implications for Further Research.....	194
<i>Impact of Sample Demographics</i>	194
<i>Generalizability of Findings</i>	196
<i>Perspectives on Conference Experiences</i>	198
Summary.....	202

Appendices

A: Research Apprenticeship Interview Guide.....	206
B: Syracuse University Study Council Teacher Evaluation Survey, 1995.....	209
C: Teacher Evaluation Conference Survey 2019.....	224
D: Dissertation Study IRB Materials.....	237
References.....	245
Curricula Vita.....	260

List of Figures:

Figure 1: Conceptual Framework—Interrelated Components of Evaluation Systems.....	18
Figure 2: Interrelated Levels of Theory.....	27
Figure 3: Map—Literature Review of Evaluation Conference Components.....	29
Figure 4: Basic Evaluation System Structure for Dual Functions and Purposes.....	30
Figure 5: Blank Q-Sort Distribution Grid.....	89
Figure 6: Mixed Methods Integration Model.....	105
Figure 7: Raw Q-Sort Data Distribution Grid.....	149

List of Tables:

Table 1: Synthesis of Literature on Observation Feedback Conferences.....	37
Table 2: QUAL-QUANT-MIXED Paradigms.....	55
Table 3: Over-view of Research Approaches.....	58
Tables 4A—4E: Summary Exploratory Factor Analyses	75-80
4A: General Approaches Evaluators Use When Conducting Conferences.....	75
4B: Approaches Evaluators Use to Involve Teacher Being Evaluated.....	76
4C: Extent of Focus by Evaluators on Specific Topics During Conferences.....	78
4D: General Approaches Evaluators Use When Conducting Conferences/ General Situations and Relationships that Affect Conducting Conferences....	79
4E: Measures of Conference Effectiveness.....	80
Table 5: Legitimation for Evaluation Feedback Conferences.....	97
Tables 6A—6L: Raw Survey Data.....	118-133
6A: Identifies and Addresses Evaluator’s Concerns.....	118
6B: Evaluator Identifies and Addresses Teacher’s Accomplishments/Strengths...	119
6C: Attention to Teacher’s Input.....	120

6D: Seeking Teacher’s Input.....	121
6E: Discussion of Teacher’s Practice.....	123
6F: Summative Ratings/Assessments.....	125
6G: Evaluator’s Fairness.....	126
6H: Evaluator’s Awareness of Teacher’s Practice.....	128
6I: Provides for Follow-up.....	129
6J: Effectiveness=Strengthening Teaching.....	131
6K: Effectiveness=Cultivating Teacher-Evaluator Relationships.....	132
6L: Effectiveness=Accountability.....	133
Table 7: Summary of Independent Factor Variables.....	136
Table 8: Summary of Dependent Factor Variables.....	137
Table 9: Correlation of Effectiveness Measures (r) with Each Other.....	137
Table 10: Factor Correlations with Effectiveness Measures.....	138
Table 11: Composite Models Predicting Evaluation Conference Effectiveness.....	141
Table 12: Pre-Factored, Raw Q-Sort Data—Most Effective/Useful Statements.....	150
Table 13: Pre-Factored, Raw Q-Sort Data—Least Effective/Useful Statements.....	151
Table 14: Factor Loadings with Defining Sorts and Factor Reliability.....	154
Table 15: Summary of Q-Sort Statements by Factor Z-Scores and Rank.....	156
Table 16: Q-Sort Factor Array—1.....	158
Table 17: Q-Sort Factor Array—2.....	162

Chapter 1

Introduction

Teacher evaluation has been a part of the educational landscape for some time, but the manner in which school organizations have implemented teacher evaluation systems varies across districts and locales (Baker, Oluwole, & Green, 2013; Kraft & Gilmour, 2017). In part, these variations come from how individual organizations make the distinction between the summative and formative purposes of teacher performance evaluations (Nevo, 2006; Peterson, 2004; Popham, 2013). In many cases, school organizations implement evaluation systems that address both purposes because of state or federal accountability requirements (Baker, et al., 2013; Champ, 2015; Lavigne, 2014; Popham, 2013). School organizations and districts fulfill accountability requirements by combining data on student growth with evidence of teaching practice from observations to assign summative ratings to individual teachers; school organizations and districts will also use the collected data from student growth and observations to gauge the performance and competency of teachers for formative purposes such as developing teacher efficacy. However, when school organizations use evaluation data from the two measurements for combined purposes addressing teacher/school improvement and organizational accountability, critical tensions emerge from differing expectations and uses of the data by administrators and teachers. When these tensions exist, all stakeholders are affected and complicate using teacher performance evaluations to show accountability or help improve teaching and learning.

Currently, New York State (NYS) school districts all must submit a teacher/principal performance evaluation plan to the NYS Education Department (NYSED) for approval by the Commissioner of Education, and then implement that plan to fulfill mandated performance

reviews for New York State’s Education Law §3012-d (2015), known as the Annual Professional Performance Review (APPR) law. The law specifically distinguishes the purposes for evaluation as “a significant factor in employment decisions...as well as teacher and principal professional development” (NYSED, 2018, pg.6). Even though districts submit plans that attend to both purposes, districts struggle with using performance evaluation data for summative and formative purposes simultaneously (Frontier & Mielke, 2016). To address continuing questions and the need for clarification, NYSED recently published a revised edition of *Guidance on New York State’s Annual Professional Performance Review for Teachers and Principals to Implement Education Law §3012-d and the Commissioner’s Regulations* (2018), known as the ‘Guidance Document’ in the field, which is meant to give thorough explanations for each purpose. However, this document has muddied the situation more than clarified it, causing critical tensions at the district level from teachers and evaluators struggling with how to separate the purposes within the context of their over-all evaluation system.

On the one side of this struggle is how districts and organizations deal with the summative purpose of evaluation that focuses on the appraisal, judgment and measurement of teachers’ performance aligned with an organization’s goals and objectives for student achievement (Danielson, 2001; Darling-Hammond, 1990; Marzano, 2012; Mette, Anderson, Nieuwenhuizen, Range, Hvidston & Doty, 2017; Natriello, 1990; Papay, 2012; Peterson, 2004; Ponticell & Zepeda, 2004; Popham, 2013; Scriven, 1995; Stronge, 2007). The APPR law outlines two measurements to be used to determine summative performance ratings: student growth scores and observation of teaching practice. To fulfill the observation part of the law, school organizations must outline in their APPR plans how they will use a framework of teaching competences, strategies and practices, in the form of rubrics aligned with NYS

Teaching Standards (2011), as the tool to appraise, judge and measure teachers' practice based on observations (Danielson, 2001; Danielson & McGreal, 2000; Lavigne, 2014; Marzano, 2017; Marzano, Frontier, & Livingston, 2011; NYSUT, 2019; Weems & Rogers, 2010). According to the law and Commissioner's regulations, these observation ratings are combined with student growth scores to determine the over-all summative teacher evaluation ratings which are then used to make human capital decisions that will impact the way the school organization addresses state/federal educational accountability requirements and implement organizational plans for teacher development that will impact student learning (Baker, et al., 2013; Champ, 2015; Hallinger, Heck, & Murphy, 2014; Halverson & Clifford, 2006; Hinchey, 2010; Lavigne, 2014; Nevo, 2006; Natriello, 1990; Papay, 2012; Peterson, 2004; Popham, 2013; Steinberg & Donaldson, 2016; Stronge, 2007). Part of the tensions that come from using evaluation data solely for a summative purpose is the implied expectation that teachers will understand, accept and use those summative ratings to improve practice even when they feel the ratings of their performance were for "the benefit of some *external* audience or decision-maker" (italics in original; Scriven, 1991, p. 340) such as policy makers and organizational leaders.

The other side of the struggle is how districts and organizations deal with the formative purpose for evaluation that focuses on the identification, support and motivation for professional development of teaching practices (Frontier & Mielke, 2016; Marzano, 2012; Mette, et al., 2017; Natriello, 1990; Papay, 2012; Peterson, 2004; Ponticell & Zepeda, 2004; Popham, 2013; Scriven, 1995; Stronge, 2007). In general, the formative approach to evaluation relies on multiple participants understanding the multiple components which create a culture of effective supervision that aims to support teachers and student learning through a diverse range of evaluative activities (Danielson, 2001; Darling-Hammond, 2014; Hallinger, et al., 2014;

Hinchey, 2010; Kraft & Gilmour, 2017; Marzano, 2012; McLaughlin, 1990; Nevo, 2006). For many organizations, “[t]raditional evaluation procedures include pre-observations and the completion of approved evaluation documents,” and “post-observation conferences remain a foundation” (Weems & Rogers, 2010, p. 22) of the over-all process as a way to support the formative purpose of conducting teacher evaluations (Danielson, 2001; Donaldson, 2013; Stronge, 2007). When using such a formative approach, the evaluation process becomes a tool that “provides feedback on teachers’ instructional strengths and weaknesses, highlights areas for improvement, and supports teachers’ continued development” (Papay, 2012, p. 124). When performance evaluation data (derived from observations and student outcomes on classwork) are set in the framework of an evaluation feedback conference, teachers and evaluators have an opportunity for formative and reflective interactions/discussions that can focus on the teachers’ practice that will make the over-all evaluation process meaningful for teachers and students.

In NYS, Education Law §3012-d specifically requires observations by ‘lead evaluators’ or ‘supervisors’ and by an ‘independent’ evaluator, who could also be a peer/teacher leader. The law and Commissioner’s regulations dictate the number of times evaluators are to observe teachers and how much weight each of the evaluator’s ratings have for the observation portion of the teachers’ summative ratings, but there are no mandated requirements for how evaluators are to conduct the follow-up conference with teachers about their observations ratings that would constitute addressing the formative purpose of the evaluation system. All other components related to the observation, including how/when/if there will be follow-up conferences or other activities that address formative purposes, are negotiated by districts/organizations and their teachers’ union. The vague way the APPR law addresses how districts attend to the ‘formative’ use of evaluation data, as opposed to the very explicit way districts must attend to the

‘summative’ use of data, may be contributing to the tension teachers and evaluators feel when trying to use over-all evaluation data for summative and formative purposes simultaneously.

When evaluation systems do not distinguish or clarify for which purpose evaluation data are being used throughout the evaluation process, a disconnection is created between the summative and formative purposes and the supervision-evaluation functions that impacts the over-all utility of the evaluation system. When this disconnection is present, teachers and some evaluators feel that evaluation ratings, determined by combining student performance data with classroom observations, cannot be used effectively for either purpose or function (Frontier & Mielke, 2016). When this tension between the summative and formative purposes exists, the evaluation system cannot function in ways that meet students’ educational needs, teachers’ and principals’ professional development needs, or the school organization’s needs for accountability. As a result, the summative-formative tension contributes to the issue of whether the function of teacher performance reviews should be about supervision or evaluation, with the field of educators and administrators debating between separating the purposes and functions or trying to somehow reconcile them within the over-all framework of a teacher evaluation system (Hallinger, et al., 2014; Halverson & Clifford, 2006; Hinchey, 2010; Kraft & Gilmour, 2017; McLaughlin, 1990; Mette, et al., 2017; Papay, 2012; Peterson, 2004; Popham, 2013).

The question of whether the teacher evaluation process can simultaneously provide a summative assessment of teaching/learning and formative support for professional development for teachers is central to the summative-formative debate. This research study explores the evaluation feedback conference as an untapped resource for teachers and evaluators to make evaluations more effective for addressing the supervision and/or evaluation function of the over-all evaluation system that is at the center of this summative-formative issue. To investigate how

organizations may approach resolving the issue, this research study asks teachers about their experiences with evaluation feedback conferences and how those conferences are conducted for addressing both purposes within the evaluation system in place.

Background on Evaluation Functions and Purposes

From a human resource position, evaluation is a systematic process of examination, investigation, and knowledge production for determining the merit, worth and value of something or someone; “evaluations are the products of that process” (Scriven, 1991, p. 1). In the case of teacher evaluation, the over-all design of the evaluation process needs to distinguish the system’s function as one that evaluates the productivity (as in process/product out-put) of an employee’s job performance or supervises personnel’s competency in the job performance (Baker, et al., 2013; Frontier & Mielke, 2016; Hinchey, 2010; Mette, et al., 2017). Educational organizations need to be mindful of how the functions of performance evaluation and personnel supervision are distinct and can cause critical tensions with personnel if the distinction about functions is unclear, or when the organization subsumes the personnel supervision function under performance evaluation to serve dual summative-formative purposes (Frontier & Mielke, 2016; Popham, 2013; Scriven, 1991). An effective evaluation system should reflect how the organization separates the two functions and communicates how the evaluations will be used for summative or formative purposes related to personnel decisions (Frontier & Mielke, 2016; Joint Committee on Standards for Educational Evaluation [JCSEE], 2009; Ponticell & Zepeda, 2004).

At one time, NYS school districts locally operationalized their evaluation systems for appraising teaching efficacy and school district management with general oversight by the state. More recently, the state has become a more intrusive entity at the individual district level with Education Law §3012-d. Education Law §3012-d mandates a broad-scale evaluation system that

assesses the teaching and learning going on across the state. Creation of one over-arching, specific personnel evaluation system for the whole of NYS directly relates to how NYSED is complying with federal accountability requirements to connect teachers' practice-performance with student performance on standardized assessments to receive federal funding (Champ, 2015; Hinchey, 2010; Lavigne, 2014; Steinberg & Donaldson, 2016).

In the most current iteration, NYS law specifically addresses how teacher/principal composite evaluation ratings will be determined using two measurements: student growth scores and observed teaching practice scores. The student growth measure comes from how students score on NYSED-designed/approved assessments, which are aligned to teachers' or districts' student learning objectives. Individual teachers are assigned a portion, roughly 50%, of their over-all evaluation rating based on the percentage of students whose scores meet or exceed the assigned student learning objective targets. Teachers' observation scores, minimally one by a building administrator *and* one by an independent/outside observer (usually an administrator within the district, but *not* assigned to the teacher's building), are based on observation rubrics focused on dimensions of teaching that are 'observable,' and the ratings from the observations contribute to the other portion of teachers' over-all evaluation rating. As stated in the law, districts and educational organizations use the combined measures to determine ratings and thus be used as a "significant factor for employment decisions," and as a "significant factor in teacher and principal development" (NYSED, 2018, pg. 6). The dual functions/purposes of employment decisions and professional development of school personnel as mandated in the law make personnel-performance evaluation a high-stakes process that has an impact on the academic welfare of students and the professional welfare of teachers (Baker, et al., 2013; DiPaola & Hoy, 2014; Frontier & Mielke, 2016; Steinberg & Donaldson, 2016).

In essence, the NYS law explicitly outlines the basic goals, structure and implementation requirements for school districts' APPR plan. The law's requirements impose a framework for specific evaluation and supervision functions (assessment of productivity and competence) through a combination of measures of student performance and direct observations of teachers' performance. School districts must create an APPR plan that meets these explicit requirements; however, the language of the law has the potential to create tension around how the school organization uses the evaluation data to fulfill both functions and purposes of the system. The law uses explicit and specific language, at length, on the methodology for determining the summative APPR ratings (Sections 5.a; 7.a-c) from student assessment scores, what constitutes as an assessment of student growth (Section 4.a) and how student growth assessments should be factored into teachers' over-all effectiveness ratings (i.e. HEDI scores). The law then uses broad generalizations in substantially shorter sections for outlining the required provisions related to teacher observations (Section 5.b) and "prohibited elements" for the observation subcomponent (Section 6). The lack of specificity and attention for determining the observation scores and how those scores contribute to the summative evaluation ratings, as the law does for including student growth scores in the summative ratings, creates an imbalance between the functions of evaluation and supervision that hinders school organizations from equally addressing summative and formative purposes within the same system. This imbalance then contributes to the perception or mixed message that the APPR system, in general, is only about the student scores and what those scores mean for teachers' effectiveness rating (Baker, et al., 2013; Mette, et al., 2017; Steinberg & Donaldson, 2016).

This imbalance distorts the lines between supervision and evaluation (Frontier & Mielke, 2016), thus creating the tension that teachers and administrators feel when individual districts

interpret and then implement the evaluation system without distinguishing the purposes or functions the evaluation data serve. A contributing factor in creating this imbalance comes from the requirement in the law that districts negotiate with the teachers' unions to choose a teaching practice-observation framework/model aligned with NYS Teaching Standards (2011) to determine effectiveness of teachers' practice. NYS provides a list of approved observation-teaching practice models/frameworks to use for determining observation scores that fulfill the requirements set forth in Education Law §3012-d. The most widely used frameworks and models on the list are designed such that the over-all scope and dimensions included in each of the frameworks/models meet the required indicators of effective practice outlined in the teaching standards. Each framework/model uses language broad enough to evaluate generic and content-specific practices, while at the same time allowing organizations and teachers to adjust the framework/model for their specific contexts (Charalambous, Komitis, Papacharalambous, & Stefanou, 2014; Mielke & Frontier, 2012). The distinction between the frameworks/models comes from whether the structure of the evaluation instrument focuses on teaching practice input (pedagogical decisions and instructional strategies) or teaching outcomes (results of student assessment) in relation to the data/evidence collected to fulfill the evaluative purposes set by the organization (Marchant, David, Rodgers, & German, 2015). As an evaluation tool, teaching practice frameworks function as the means to identify the content of teachers' practice, and school organizations need to consider their goals, priorities, and culture/context when making decisions about which framework to include in the design of the system.

On the one hand, this flexibility to choose allows districts and teachers' unions to include elements/dimensions of teaching in the observation rubric/protocols that would reflect an authentic assessment of teaching practices that best fit the individual district's culture and

context. On the other hand, this flexibility creates variations between districts when applying different teaching models/frameworks that use different rubrics/protocols, which may distort data for comparison purposes related to instructional practices that have an impact on student achievement or for how teachers show evidence of meeting teaching standards. The result, then, are critical tensions within and across school organizations that come from how administrators-evaluators and teachers use and interpret the evaluation process for the separate functions/purposes while also fulfilling the requirements of NYS Education Law §3012-d (Conley & Glasman, 2008; DiPaola & Hoy, 2014; Donaldson, 2016; Donaldson & Donaldson, 2012; Frontier & Mielke, 2016; Halverson & Clifford, 2006; Herlihy, Karger, Pollard, Hill, Kraft, Williams, & Howard 2014; Lavigne, 2014; Marzano, et al., 2011; Natriello, 1990; Nevo, 2006; Peterson, 2004; Popham, 2013; Reid, 2017; Scriven, 1991).

One way some educational organizations have attempted to lessen the tension between the functions/purposes is to use the observation/evaluation feedback conference as an activity as part of the over-all evaluation process. During this feedback conference, the evaluator and teacher can discuss, review and address how the teacher demonstrates evidence of highly effective teaching practices as well as discuss the data from student work/assessment scores using the evidence gathered by the evaluator. The discussions, then, become an integral part of the evaluation process when included as part of the over-all evaluation system (Behrstock-Sherratt, Rizzolo, Laine, & Friedman, 2013; Danielson, 2009, 2015; Hall, 2019; Helm & St. Maurice, 2006; Hopkins, 2016; Marzano, 2017; MET Project, 2015; Popham, 2013; Tuytens & Devos, 2017). As another negotiated component of the APPR plan, districts/organizations and their teachers' unions have the opportunity to craft how the evaluators will conduct feedback

conferences for these formative discussions that specifically address their specific needs for improving teaching and learning.

In theory and sometimes in practice, the observation feedback conference is a key activity in the evaluation process when an evaluator and teacher can unpack practice and student learning by having a two-way conversation (Hall, 2019). The conversation brings the teacher's perspective on practice and the evaluator's observation of practice to the center of the feedback conference to realize the greatest potential to change what happens in the teachers' day-to-day instructional context (Arneson, 2015; Danielson, 2015; Hill & Grossman, 2013; Kise, 2014; MET Project, 2015; Reilly, 2015; Renfro, 2014; Roussin & Zimmerman, 2014; Tuytens & Devos, 2012, 2014). The feedback conversation should provide time and opportunity for deep, pedagogical discussions about student and teacher learning for the intent of having an impact on a teacher's practice that in turn has an impact on students (Myung & Martinez, 2013; Popham, 2013; Tuytens & Devos, 2014). However, the feedback conference also has the potential to create a dichotomy between teachers and evaluators (Calabrese, Sherwood, Fast, & Womack, 2004) over the summative and formative purposes of performance ratings if the feedback discussion focuses on one purpose at the expense of the other. When teachers and evaluators do not have a clear consensus about the use of evaluation data while engaged in the feedback conference, the tension around the evaluation's function and purpose create a disconnection ripple effect that can negatively impact the efficacy of the evaluation process over-all.

Problem Statement

The persistent argument in the field is whether or not teacher evaluations, especially feedback conferences, can serve both the summative and formative purposes within a single evaluation system. Even when school districts were developing evaluation systems without the

state/federal accountability oversight or policy requirements, there were issues with how educational leadership (i.e. administrators) would approach designing, implementing and using the assessment of teachers' performance (Hall, 2019; Mette, et al., 2017). This current climate of accountability has exacerbated the tension between the summative-formative purposes, and in turn has started to erode the impact of teacher evaluation over-all.

If accountability-based evaluation systems continue to focus principals' and teachers' attention on complying with steps established by law, and if fulfilling the steps continues to be more important than the process of adult learning required to improve teaching and learning, then there is little hope that supervision and evaluation will be perceived by teachers or principals as anything more than a perfunctory, compliance-centered process where both principals and teachers deliver the required show (Ponticell & Zepeda, 2004, pg. 54).

When one purpose or function seems to undermine or diminish the potential of the other, the result is neither purpose or function will be able to have an effect on student learning or teacher development (Conley & Glasman, 2008; Darling-Hammond, 2014; Donaldson, 2013, 2016; Donaldson & Donaldson, 2012; Halverson & Clifford, 2006; Papay, 2012; Popham, 2013).

The more recent attention on how teacher evaluations contribute to how school organizations fulfill state and federal accountability requirements/mandates also has drawn attention to how evaluations are conducted. There seems to be a perception that many school organizations' evaluation systems are flawed for how they incorporate and balance the dual functions/purposes of supervision and evaluation (Darling-Hammond, 2013, 2014; Frontier & Mielke, 2016; Hall, 2019; Popham, 2013) which then impedes the efficacy of evaluation systems to have the impact on teaching and learning that they are meant to have. Despite the extensive body of research and literature on the differentiation between supervision and evaluation, which includes the debate on the summative-formative purposes, there is no definitive resolution for how to accomplish such a balance. The purpose of this study is not an attempt at finding or

forming definitive answers, but endeavors to explore how, if possible, school organizations and teachers can resolve or reconcile issues and tensions related to the way school organizations conduct evaluative activities such as feedback conferences to address dual purposes simultaneously within the context of evaluation systems.

This study will also shed light on an aspect of instructional leadership experienced by administrators and/or evaluators who struggle with how to respond to performance issues that result from evaluations and the APPR process within the context of the evaluation feedback conference (Hall, 2019). By understanding how teachers and evaluators conduct these conferences within a constructed evaluation system, educational organizations will have a resource that would support developing a framework for evaluation conferences that will be effective for all participants and provide actionable insights that would (re)engage both teachers and administrators in evaluative endeavors that would promote powerful impacts on teaching and learning.

Research Questions and Rationale

The tension at the center of the summative-formative argument poses questions about how teachers and administrators-evaluators conduct and participate in feedback conferences that simultaneously address the dual functions and purposes of an evaluation system. Conducting a study from the teachers' perspective opens the "black box" (Halverson & Clifford, 2006; Muñoz, Scoskie & French, 2013) on how teachers feel about feedback based on observation of practice and the impact that feedback is/is not having on their practice. Examining this issue from the teachers' perspective raises a number of questions which are not addressed in the current research literature: Is it possible to reconcile the two purposes through feedback conferences as part of the over-all evaluation system? Does the system, in terms of state mandated components

and efficacy scales (i.e. HEDI), have an impact on how teachers receive/act on feedback given within the context of such evaluation conferences? Do teachers perceive the efficacy of feedback conferences with a summative or formative lens, and how does that perception impact their practice? Do teachers report differences across experience levels for being receptive to feedback given? Are certain strategies or approaches that teachers report their evaluators use for conducting the feedback conference correlated with how they perceive the conference addresses either or both purposes? Do teachers perceive the evaluation feedback conference as a way to develop their reflective skills or do they consider the feedback conference as a cursory interaction with little to no impact? What are the impacts on teacher-evaluator relationships when evaluation feedback conferences are structured within the current context of the NYS APPR evaluation system, as teachers' responses to the survey items indicate? A research study on utility and efficacy of evaluation feedback would focus on answering these questions related to the experiences of teachers with performance-feedback conferences, and how the conferencing-feedback skills of lead evaluators have an impact on teachers taking ownership of their evaluations for accountability (summative purpose) and professional learning/improvement (formative purpose).

The responses collected during this study will provide insight on whether teachers consider particular approaches and strategies necessary for evaluation conferences to be effective for addressing summative and/or formative purposes. A potential outcome for the study is to provide organizations with guiding factors to consider when developing a framework for training evaluators to conduct effective feedback conferences based on information from teachers' responses. By developing a better framework for conducting effective feedback conferences, teachers and educational leaders could address the tensions, with the goal to lessen if not

eliminate them, that come from using over-all performance appraisals summatively and formatively within the same evaluation system (Halverson & Clifford, 2006). With this outcome in mind, this study focuses on three research questions:

1. What are teachers' experiences with how evaluators conduct evaluation feedback conferences?
2. Is there a connection between the way evaluators conduct feedback conferences, as reported by teachers, and how those conferences affect teachers' practice?
3. Based on teachers' reports, under what circumstances (if any) is it possible for evaluation feedback conferences to serve both summative and formative purposes?

The first question focuses on teachers' lived experiences from their perspectives, collective knowledge and sense-making of the evaluation feedback conference experience. The second question explores whether and how teachers perceive evaluators provide feedback during the conference that is useful and actionable. The third question investigates the possibility of the feedback conference being the key to reconciling the tensions participants feel in the evaluation process when the process serves two purposes.

Theoretical Framework

An integral portion of any research study's design and plan is how to best choose the research approach that will capture all the complexities of the topic under study. The theories connected to that approach are equally as important as the approach itself. The design for a study of teachers' experiences with evaluation feedback conferences needs multiple theories and research approaches because there is no consensus on a singular theory of teaching or evaluation of teaching, based on a mono-methodological approach, that would adequately capture the complex nature of evaluating teachers' practice (Chambers, 1992). A pragmatic, mixed methods approach is best suited for this study since the mixed methods researcher accepts multiple

ontological, epistemological, axiological and methodological assumptions and frameworks and avoids narrowly categorizing the collected data to fit a priori theoretical frameworks or assumptions (Beuving & de Vries, 2015; Chambers, 1992; Creswell, 2013; Lincoln & Guba, 1985; Long & Rodgers, 2017; Patton, 2015).

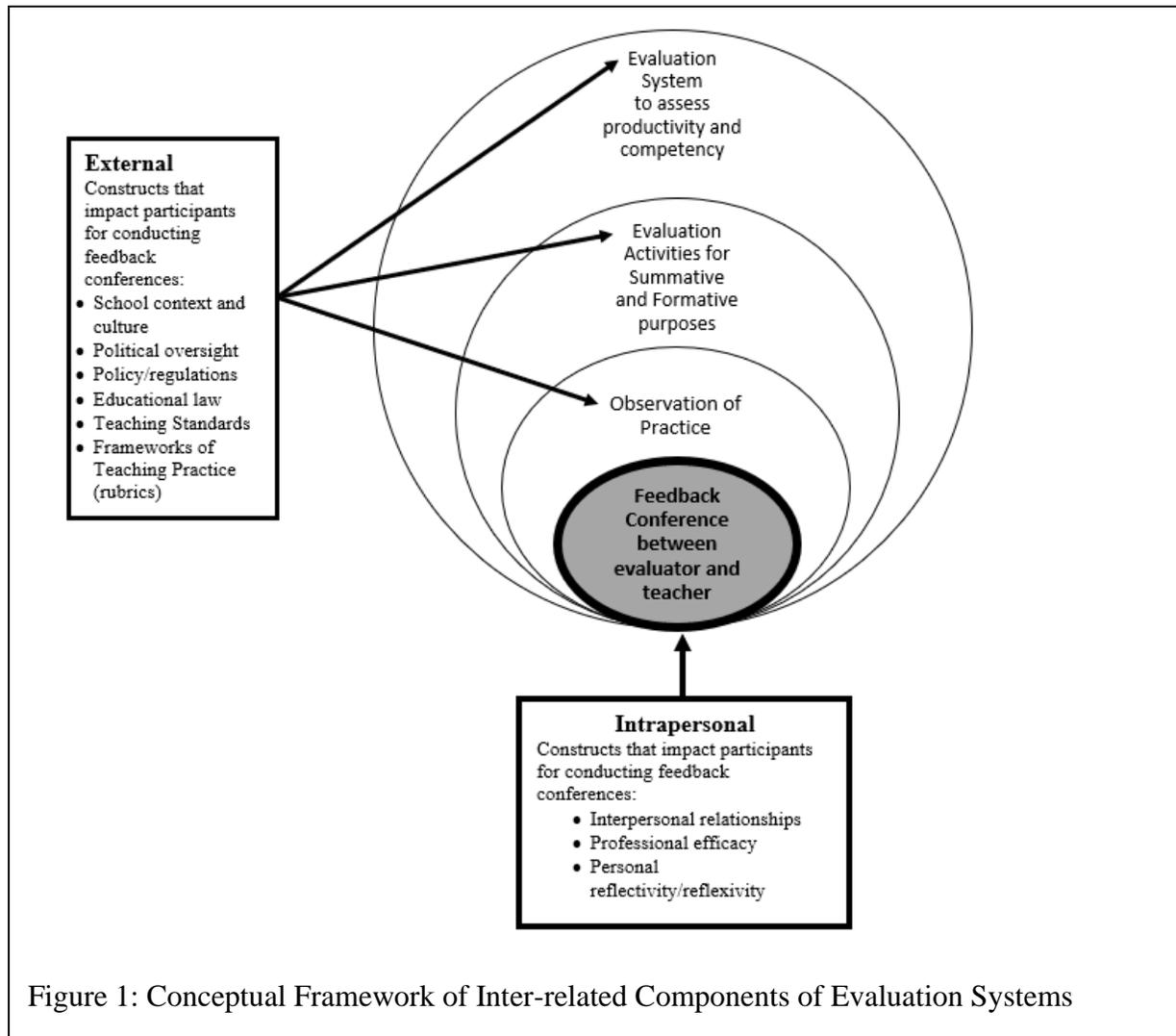
By approaching the study pragmatically, the researcher “puts methodological theory at its core and includes the explicit articulation of the relationship between theoretical and practical aspects of educational research” (Long & Rodgers, 2017, p. 2813). When methodological theory is put in the center of the study, rather than an a priori theory of the phenomenon being investigated, the interpretations of the research findings make the over-all study fit the paradigm of pragmatic applied research (Hedrick, Bickman, & Rog, 1993). That is particularly appropriate when, as in this study, different theories purport to explain parts of the same phenomenon but not whether or how they might be reconciled. Using methodological theory to focus the study supports application of findings from the individual parts of the research process. Each level of theory informs the other levels, with the result being each level shapes the way the researcher collects and interprets the data at multiple stages of the study.

Current discussions related to incorporating multiple philosophical assumptions and theoretical frameworks acknowledge that various socio-behavioral, political and cultural issues defy the notion of fitting only one specific tradition (Beuving & de Vries, 2015; Cameron, 2011; Creswell, 2013, 2014; Leech & Onwuegbuzie, 2009; Long & Rodgers, 2017; Patton, 2015; Strega, 2015; Teddlie & Tashakkori, 2012). Therefore, as part of the design process for a mixed methods study, the researcher must consider how philosophical assumptions guide the use of theoretical frameworks, and then consider how the frameworks have various beliefs and assumptions embedded within them (Cameron, 2011; Chambers, 1992; Creswell, 2013; Leech &

Onwuegbuzie, 2009; Long & Rodgers, 2017; Patton, 2015). The assumptions are then woven with the theoretical frameworks to become the interpretative lens that guides the researcher throughout the implementation of the study (Beuving & de Vries, 2015; Bogdan & Biklen, 2007; Creswell, 2013; Denzin & Lincoln, 2000; Lincoln & Guba, 1985; Long & Rodgers, 2017; Onwuegbuzie & Leech, 2005; Patton, 2015; Teddlie & Tashakkori, 2012). Ultimately, the researcher's choice or preference of philosophical/theoretical framework resides in the topic of the inquiry, researcher subjectivity, and holistic use of the findings rooted in a specific discipline and core questions guiding the study (Chambers, 1992). Further discussion of the inter-related theories for this study is included as the introduction to Chapter 2: Literature Review.

Conceptual Framework

The conceptual framework for this study is a synthesis of micro-substantive theories on *feedback* (Ilgen, Fisher, & Taylor 1979; Kinicki, Prussia, Wu, & McKee-Ryan 2004; Kluger & DeNisi, 1996), *educational performance evaluation* (JCSEE, 2009; Popham, 2013; Scriven, 1991, 1995), and *performance feedback conferences* from the field of clinical supervision (Danielson & McGreal, 2000; Danielson, 2015; Marzano & Toth, 2013; Marzano, et al., 2011), mentoring (Behrstock-Sherratt, et al., 2013; Costa & Garmston, 2002; Lipton, Wellman, & Humbard, 2003), and educational supervision/evaluation (DiPaola & Hoy, 2014; Frontier & Mielke, 2016; Glickman, 2002; JCSEE, 2009; Lipton & Wellman, 2013; Marshall, 2013). Figure 1 represents the complexity of an evaluation system that includes multiple components that address dual functions and purposes:



Each component of the system is nested within a larger component, with the outer-most one representing the over-arching conceptualization of what an evaluation system should be used for in an organization. Each smaller embedded component then represents how the system moves from abstract function to concrete action, with the inner-most ring indicating that the experience between a teacher and someone who observes that teacher's practice, being a school district leader, designated administrator—supervisor or lead evaluator, sits at the core of the system. This feedback conference may be just one component of the over-all system, but the way the

conference is embedded in the process contributes to the tension that complicates how participants engage in the feedback conference.

The conceptual framework also indicates external and intrapersonal factors impact how teachers and administrators/evaluators engage in that one activity at each stage of the evaluation process. The external factors impact the over-all system (indicated by the three outer component circles) for how the evaluation process is implemented. These external factors manifest as how the evaluation system has been structured to conform to mandates, policies, and requirements independent of the actors most directly involved with implementing such a system, namely administrator-evaluators and teachers. The intrapersonal factors reside within/between teachers and administrators/evaluators, and manifest as teachers' internalized thinking or outward communicative expression for how the experience impacts their practice. The impact of intrapersonal factors is more difficult to ascertain since each participant in the conference, teacher and/or administrator-evaluator, uses a different affective lens for making sense of the situation; generally, how teachers affectively respond to and are receptive of what happens during a conference experience may be from a different affective position than how evaluators affectively respond to and are receptive of what happens during the conference. On the surface, the external and intrapersonal factors seem to independently impact different components of the evaluation system, where external factors undergird the function, purpose and activities of the process, and intrapersonal factors highlight the subjective reactions to the function, purpose and activities of the process. In actuality, these factors are intertwined at all stages of implementation, contributing to the complexity of conducting feedback conferences when districts and school organizations simultaneously address dual functions and purposes in the same system.

Methodology

This study of how teachers make meaning of feedback within the context of the evaluation conference uses a mixed methods research design, focusing on data collection that make experiences visible in order to be understood by the researcher (Beuving & de Vries, 2015; Bogdan & Biklen, 2007; Denzin & Lincoln, 2000; Onwuegbuzie & Leech, 2005). As a general definition, a study that follows a mixed-methods model “represents research that involves collecting, analyzing, and interpreting quantitative and qualitative data in a single study or in a series of studies that investigate the same underlying phenomenon” (Leech & Onwuegbuzie, 2009, p. 267). Following such a design means focusing on describing, analyzing and reflecting on the patterns which emerge from collected empirical data (qualitative and quantitative) in order to understand not just the phenomenon, but the impact the social, political, and cultural contexts have on the experiences related to the phenomenon (Cameron, 2011; Chambers, 1992; Leech & Onwuegbuzie, 2009; Long & Rodgers, 2017; Onwuegbuzie & Leech, 2005; Saldaña, 2013; Tashakkori & Teddlie, 2010; Teddlie & Tashakkori, 2012). Specifically, this study combines a qualitative study using interviews with a quantitative study using survey research, which culminates with using a mixed method approach (Q-Methodology) with focus groups to investigate the evaluation feedback conference phenomenon. Chapter 3 provides the framework for each approach more fully.

Delimitation

A complete evaluation system that aligns with NYS APPR regulations includes multiple measures of teachers’ effectiveness, namely student growth/performance scores and observations of classroom practice. The combination of these measures results in the over-all evaluation ratings of teachers’ practice (i.e. HEDI ratings). Because there is extensive literature and research

studies that have explored the value and impact of the student performance measure on over-all teacher evaluation ratings, this study acknowledges student growth measures and assessment scores by including brief survey items as one of the topics for evaluators to discuss with teachers as part of the entire evaluation process and feedback conference context. This study does not explore teachers' perceptions on the student performance portion of their evaluation ratings in depth.

Definitions of Terms

The following are terms specific to the research study and will be used for the reporting of the findings. In general, the terms are defined as they would be used in the field. As such, the definitions are summaries of the concepts used across multiple references related to the literature on evaluation systems and design, supervision of teaching, and educational leadership (see "References" list).

Accountability: A broad term for taking responsibility for practices and student achievement data. There are two distinctions to be made for the term as used in this study:

- Accountability for *educational organizations* refers to how the district/organization will provide measurable data (i.e. summative evaluation ratings, student scores on assessments, graduation rates) to meet progress-achievement targets set by state and federal regulations for funding or oversight status (i.e. schools identified as low-performing or exceptional).
- Accountability for *teachers/evaluators* refers to reflection on/acceptance of evidence for how instructional practices impact student learning or reflect what are thought to be appropriate approaches to teaching. Based on the evidence, teachers/evaluators plan for changes/continuation of practices to show accountability as part of the evaluation process/cycle.

Administrator(s): A member of the educational organization that has a leadership role within the organization. This role includes, but not limited to, supervising, evaluating, and/or making employment decisions based on evaluations for teachers within the organization.

Affective Constructs: The subjective, intrinsic elements that impact how evaluators and evaluands use, respond to, make meaning of, or reflect on information as a result of communicative experiences during evaluation activities.

Affective Reception: A subjective, intrinsic reaction for when a person receives feedback, which prompts an extrinsic response that will reflect how the person accepts the feedback (i.e. feedback given viewed as respectful, useful, accurate, authentic, fair or opposites).

Affective Responses: The subjective, intrinsic and internalized emotions, beliefs and attitudes teachers personally and individually have and/or hold that reflects the professional identity they hold of themselves and their work within that professional organization.

Authentic/Authenticity: In general, either term is used when referring to what is perceived or described as something (i.e. rating of practice, evaluator's approach or comment) that shows a genuine intention and/or sincerity for collaboration/appraisal, implying trustworthiness as well.

Communicative Experience: The interaction between evaluators and evaluands, where each participant brings forward specific topics, concerns or issues for discussion, that is at the center of the observation feedback conference evaluation activity.

Competency: A teacher's outward projection in practice which shows his/her knowledge, capacity and ability to engage students/others for the purpose of instruction and learning.

Conference: The meeting between an administrator/evaluator with teaching personnel for the purpose of discussing the evidence collected of teaching performance as required by the organization's evaluation system.

Constructed Evaluation Experience: The interactions of evaluators and evaluands that relate to the evaluation system as a whole; experiences are framed by the external construct as mandated by educational policy or law.

Educational Organization(s): The collective resources, including all stakeholders, that support the educational endeavors of a specific/distinct area or school district.

Effective(ness): A broad, descriptive term(s) applied to particular aspects of the process and/or effects of the evaluation system, including evaluators' approach (process) to feedback conferences for having an impact on teachers' practice/accepting responsibility for the evaluation (effect).

Evaluand: The person, specifically a teacher, who is evaluated according to the school organization's constructed evaluation system.

Evaluation: The cognitive and practical process of determining the merit, worth and value of teaching for the purpose of making a judgment that has implications for making human capital decisions and building capacity within the ranks of the school personnel for effective teaching practices.

Evaluator(s): The person or persons, usually an administrator within the school organization, with the responsibility to complete the process of the evaluation of school personnel.

Evaluation System: The design, procedures, instruments (i.e. rubrics of teaching practices, state mandated efficacy scales, assessment requirements), and protocols that are combined into an analytical process for determining merit, worth and value of the school personnel being evaluated.

External Constructs: The external mandates, perceptions or policies related to teacher evaluation systems that impact how evaluators and evaluands use, respond to, make meaning of, or reflect on information as a result of communicative experiences during evaluation activities.

Feedback: A broad term that includes the oral as well as written commentary on a teacher's performance, usually (but not limited to) for the purposes of encouraging a change in practices, highlighting issues or concerns about practices or student achievement, and/or maintaining current professional progress that will meet the school organization's goals and objectives.

Formative: A broad, descriptive term used to indicate the purpose or function of an activity (i.e. assessing student learning, giving feedback on teaching practice, collaborating with others towards an organization's goal/objective for practice) as mainly for developing the necessary skills or knowledge of the personnel involved that would increase the efficacy of performing the task.

Productivity: The external connection made between student scores/output and teachers' instructional practices linked to student scores/output; external means to assign efficacy rating for teachers by the educational organization.

Professional Development: A broad term that is applied to cognitive and practical activities related to increasing awareness of and ability with various instructional concepts that will have an impact on the pedagogical/methodological approach of teachers and student learning in general.

Ratings: The numerical score for teacher performance that results from evaluators using rubrics for assessing teaching performance and a scale for student achievement on assessments.

Reflective(ity): The intrinsic ability to objectively evaluate and articulate how actions and decisions impact practice.

Reflexive(ity): The intrinsic ability to consider and articulate how decisions and actions impact self and practice, both objectively and subjectively.

Summative: A broad, descriptive term used to indicate the purpose or function of an activity as mainly being the cumulative step for making a judgment, an over-all appraisal or a decision about the level of proficiency in performing a task.

Supervision: The act of creating conditions for professional growth and development through support and cognitive coaching that empowers teachers' self-awareness, reflection and reflexive skills.

Summary

Many (perhaps even most) teachers and educational leaders in the field would say that there is no—or little—impact on teachers’ practice from any part of the performance evaluation process or from any feedback suggestions or ratings of lead evaluators (Calabrese, et al., 2004; Frontier & Mielke, 2016). This study questions whether this conventional wisdom/perception is accurate across the board or whether there are exceptions from which policy leaders and practitioners can learn. The investigation of teachers’ perceptions could possibly answer the questions related to how the summative purpose has seemed to out-weigh the formative professional development opportunity that the process presents, especially regarding how observation feedback conferences are conducted as part of the over-all evaluation process. Further investigation could also provide insights to how evaluators and teachers are able to balance the two purposes during or with observation feedback conferences based on positive, collaborative discussion and mutual understanding of effective teaching that would have an impact on practice and student learning.

Chapter 2

Literature Review

The over-all purpose of this study is to explore the topic of teachers' perceptions of evaluation conference experiences for how evaluators' approaches may/may not impact the effectiveness of the conference activity. The following literature review begins with a brief explanation of how specific macro-metatheories and micro-substantive theories inform the research design, instrumentation, implementation and data collection for this study. Following the theoretical literature review is an accounting of the practical literature on teacher evaluation systems and implementation, and discussion of current research on teacher evaluation from the vantage of teachers' perceptions.

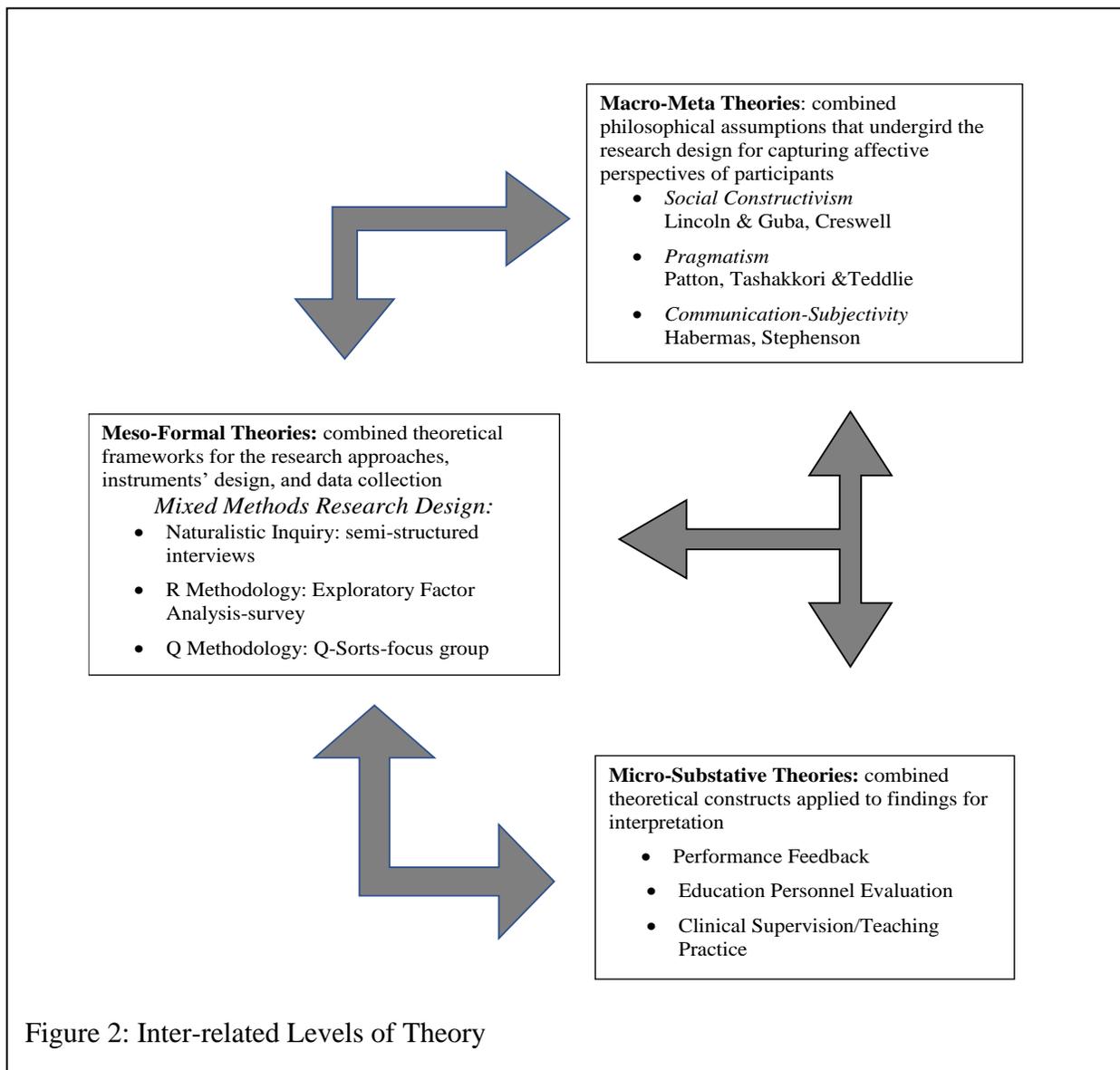
Literature on Theoretical Frameworks

The study of teacher evaluation function, design, and procedures bring together multiple ontological, epistemological, axiological and methodological assumptions and frameworks characteristic of mixed methods research. Research methodologists note that this underlying characteristic avoids narrowly categorizing the collected data to fit a priori theoretical frameworks or assumptions (Beuving & de Vries, 2015; Creswell, 2013; Lincoln & Guba, 1985; Long & Rodgers, 2017; Patton, 2015), and positions methodological theory at the center of the study for "the explicit articulation of the relationship between theoretical and practical aspects of educational research" (Long & Rodgers, 2017, p. 2813). When methodological theory is put in the center of a mixed methods study, other than an a priori theory of the phenomenon being investigated, the interpretations of the research findings make the over-all study fit the theoretical and practical paradigm of pragmatic applied research (Hedrick, et al., 1993).

I am using this theoretical-methodological framework for how mixed methods of collecting qualitative and quantitative data allow experiences (i.e. data) to become visible in order to be understood by the researcher (Beuving & de Vries, 2015; Bogdan & Biklen, 2007; Denzin & Lincoln, 2000; Onwuegbuzie & Leech, 2005). To apply mixed methods with fidelity, there should be a methodological focus on describing, analyzing and reflecting on the patterns which emerge from collected empirical data (qualitative and quantitative) in order to understand not just the phenomenon, but the impact the social, political, and cultural contexts have on the experiences related to the phenomenon (Cameron, 2011; Chambers, 1992; Leech & Onwuegbuzie, 2009; Long & Rodgers, 2017; Onwuegbuzie & Leech, 2005; Tashakkori & Teddlie, 2010; Teddlie & Tashakkori, 2012). In this way, the approaches associated with both qualitative and quantitative research paradigms contribute to fully and pragmatically exploring a complex issue with a “bi-focal lens” (Onwuegbuzie & Leech, 2005, p. 383). Such a lens builds the emic-insider and etic-outsider views of the phenomenon into the research study’s design, balancing the viewpoints of the researcher and the participants, that will lend legitimacy (Onwuegbuzie, Johnson, & Collins, 2011) for using mixed methods to understand the phenomenon under study.

A pragmatic mixed methods research study relies on mixed methods models (Kim, 2016; Long & Rodgers, 2017; McKeown & Thomas, 2014; Onwuegbuzie & Leech, 2005; Teddlie & Tashakkori, 2012) for framing the multiple macro-metatheories and micro-substantive theories, from both qualitative and quantitative paradigms, to guide research decisions and interpretations for practical and actionable answers to research questions. Specifically, I looked to macro-metatheories focused on social constructivism, pragmatism and communication/subjectivity as the foundational philosophical assumptions that underlie the research topic. When combined in a

mixed methods study, the macro-metatheories provide a philosophical framework for understanding how each participant brings his or her own meaning-making lens to the phenomenon framed by the micro-substantive theories that provide the practical constructs applied to the data for interpretation. The following figure shows how each level of theory contributes to the over-all design, implementation and interpretation of this study:

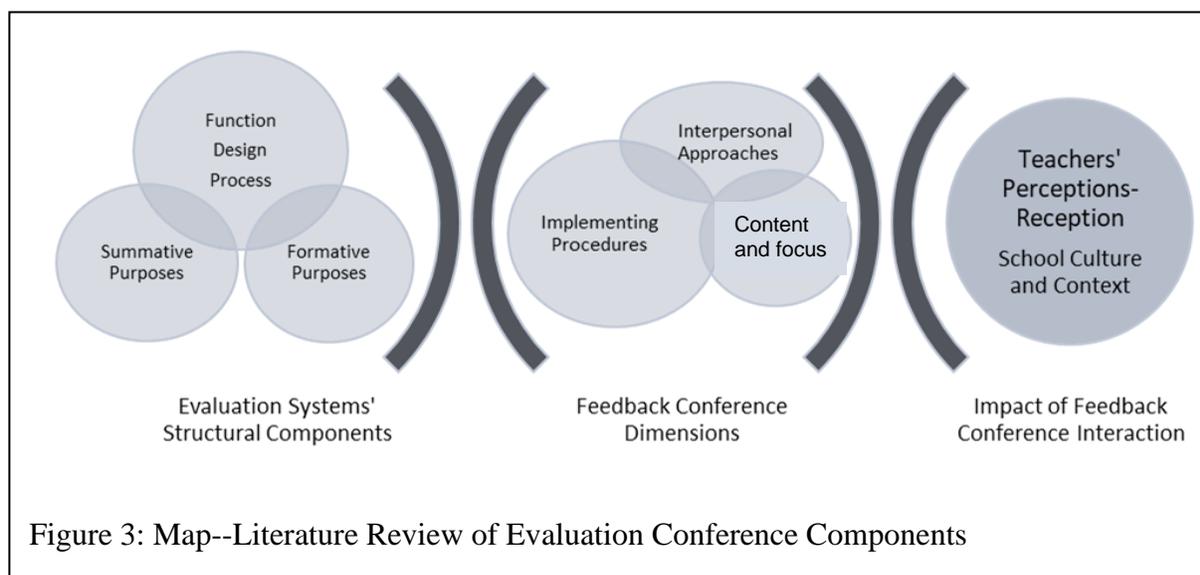


I used macro-metatheoretical frameworks related to Naturalistic Inquiry, Critical Social Constructivism, and Theory of Subjectivity (Beuving & de Vries, 2015; Creswell, 2013; Kincheloe, 1993, 1997, 2006; Kincheloe & McLaren, 2000; Lincoln, 1997; Lincoln & Guba, 1985, 2000; Patton, 2015; Phillips, 1997; Schwandt, 2000; Stephenson, 1977, 2014; Tedlock, 2000; Vidich & Lyman, 2000) since the philosophical assumptions focus on knowledge construction and individual ‘sense-making’ of lived experiences. I looked to multiple micro-substantive theories to “provide a generalized way of thinking about the major ideas and concepts discussed at the level of the metatheory” and “connect theory to the real-world application[s]” (Long & Rodgers, 2017, p. 2819). The relevant frameworks at this level include practical theories on *feedback* (Ilgen, et al, 1979; Kinicki, et al., 2004; Kluger & DeNisi, 1996, 1998), *educational performance evaluation* (JCSEE, 2009; Popham, 2013; Scriven, 1991, 1995), and *performance feedback conferences* from the field of clinical supervision (Danielson & McGreal, 2000; Danielson, 2015; Marzano & Toth, 2013; Marzano, et al., 2011), mentoring (Behrstock-Sherratt, et al., 2013; Costa & Garmston, 2002; Lipton, et al., 2003), and educational supervision/ evaluation (DiPaola & Hoy, 2014; Frontier & Mielke, 2016; Glickman, 2002; JCSEE, 2009; Lipton & Wellman, 2013; Marshall, 2013). These theories outline the inherent complexity of performance feedback within the context of a constructed experience, and how those experiences have shaped the perceptions of participants for what comprises effective observation-evaluation conferences.

Practical Literature: Evaluation Feedback Conference

The body of research literature on educational evaluation systems, including the dimensions for effectively implementing those systems, “represent[s] a national and international consensus of what is most important to sound personnel evaluations” (JCSEE, 2009, p. 1). To

support this consensus, the literature includes personnel evaluation standards specifically for education (JCSEE, 2009) and system frameworks (Danielson & McGreal, 2000; Marzano, 2017; Marzano, et al., 2011; NYSUT, 2014) which can help organizations develop the appropriate system that best fits their evaluation situation. Whether bound to developing a plan dictated by educational law, such as New York State, or locally negotiated agreements, educational organizations generally follow a similar structure that includes elemental components found in the literature and research. However, as the literature review map indicates, the literature tends to separate the specific components that make up the evaluation process from the specific dimensions of feedback conferencing and the impact of that interaction as a result of that evaluation feedback conference:

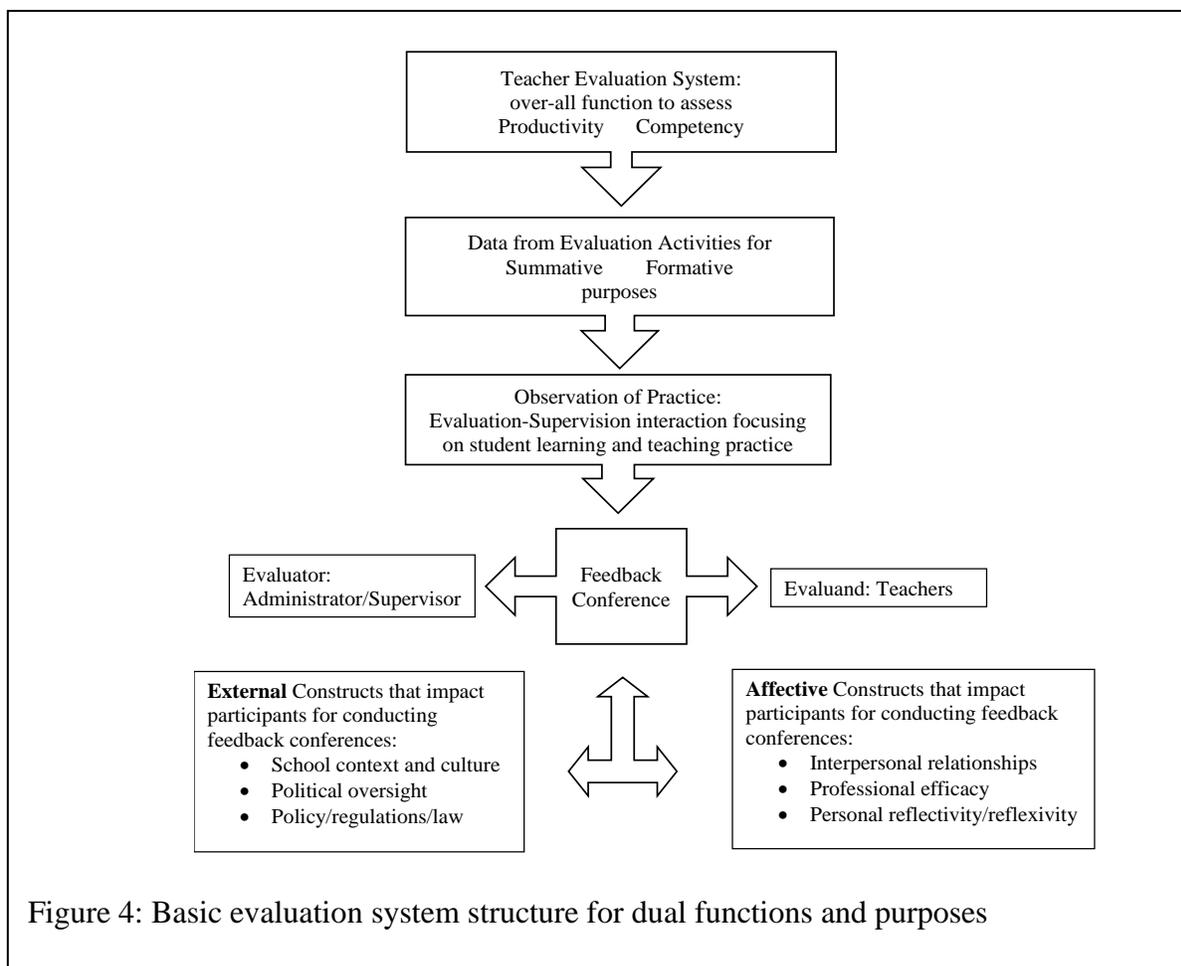


This part of the literature review will focus on the inter-related nature of these separate components of the over-all evaluation system, including how feedback conferences are an integral activity in the evaluation process. The review provides an overview of the research studies on evaluation systems' function and design, including a discussion of how summative and formative purposes intersect with other organizational considerations such as organizational

culture and context. The next section examines the literature on the varied dimensions of feedback conferences as one distinct component of an evaluation system, including the evaluators' approach, implementation of procedures, and content/focus for conducting the feedback conference. The last section focuses on existing research studies and literature on understanding teachers' perceptions and receptions of the feedback conference experience.

Teacher Evaluation Systems' Function, Design and Structure

The literature on evaluation systems' function, design and structure provides essential components that school organizations should consider for the development and implementation of an evaluation system, including procedures to ensure evaluative activities such as evaluation feedback conferences are effective for teachers and the organization.



Scriven's (1991, 1995) studies cite basic considerations that address how organizations should apply these components in the design of evaluation systems. From the start, evaluation systems need to distinguish the system's function as one that evaluates the performance-productivity of an employee (as in process/product out-put) or the job-related skills the employee possesses (as in personnel's competency) in the performance of the job (Frontier & Mielke, 2016; Scriven, 1991). The design of an effective evaluation system should reflect how the organization separates the two functions and communicates how the collected evaluation data will be used for formative or summative purposes (JCSEE, 2009; Scriven, 1991). The appraisals and judgments about work done for personnel-performance evaluation purposes should be predicated on the organization's goals, objectives and priorities to increase over-all productivity. In this sense, there needs to be a balance between the evaluation of performance function with the supervision of practice function (Frontier & Mielke, 2016; Popham, 2013). A clear distinction of what function the over-all evaluation serves becomes the foundation for the whole evaluation system's implementation and for understanding how the ratings are evidence of employees' value for the organization.

Undergirding any evaluation endeavor is attention to the ethics and validity of the process while conducting evaluations that seek to understand, appraise, and make judgments of employees' work performance, such as evaluation of educational personnel, for the value to the organization (JCSEE, 2009; Popham, 2013; Scriven, 1991, 1995). Any evaluation system can be subject to confusion over what constitutes personnel's merit-worth-value, improper evaluative practices, or unethical conduct by evaluators without clarity of purpose or function (Popham, 2013; Scriven, 1991). To ensure that education personnel are being evaluated fairly and equitably, organization administrators have a set of standards available that would assure

evaluations are conducted “in ways that are productive, valued by the profession, and that produce sound evaluative findings for serving student learning in schools” (JCSEE, 2009, p. xix). The Joint Committee on Standards for Educational Evaluation (JCSEE) developed the standards for personnel working in educational organizations to ensure “the systematic assessment of a person’s performance and/or qualifications in relation to a professional role and some specified and defensible institutional purpose” (JCSEE, 2009, p. 3). The developers of *The Personnel Evaluation Standards* considered multiple sources and research studies before articulating and defining the attributes of responsible evaluation procedures (Howard & Sanders, 2006; JCSEE, 2009; Stufflebeam & Sanders, 1990); the JCSEE *Personnel Evaluation Standards* (2009) provide a model of evaluation from a “pluralistic view regarding the application of [the] standards” (JCSEE, 2009, p. 4) so education organizations can apply them to their specific circumstances and contexts.

The JCSEE constructed the standards around four attributes of effective evaluation so that organizations could make the system “an integral part of societal and institutional efforts to prepare, engage and develop educational personnel” (Stufflebeam & Sanders, 1990, p. 4). These standards of *propriety, utility, feasibility, and accuracy* emphasize the “fundamental purpose” of evaluations is “to help provide effective services to students ... allow educators to determine the quality of how they perform the responsibilities of their work and to gain direction for improving their performance” (JCSEE, 2009, p. 1). Educational organizations are not required to apply the JCSEE *Standards* to the various systems in place, but to do so would “hold teachers accountable for the high standards demanded by the public, resulting in instruction that best benefits” students (Howard & Sanders, 2006, p. 68). Evaluation systems that use the JCSEE *Personnel Evaluation Standards* (2009) would be supporting the educational organization’s commitment to

ensuring consistency in application of the evaluation system, equity of methods to collect data, balance across purposes and functions, and constructive use of the information on performance that is free of biases or demoralizing evaluative practices (JCSEE, 2009; Stufflebeam & Sanders, 1990).

The design and structure of an effective evaluation system should combine understanding the basic principles or attributes of evaluation systems with attentiveness to the complexity of the evaluative context. Personnel-performance evaluations in an educational setting have their own levels of complexity based on the organizational culture and sociological variables or contexts (Lortie, 1975). Evaluation systems, especially in an educational context, become a way of communicating the norms of an organization that creates the culture (Darling-Hammond, 1990; Lavigne, 2014; Popham, 2013; Scriven, 1995; Tuytens & Devos, 2014). Culture and context, as related to educational organizations, derive from the collective experiences and inter-intrapersonal relationships of organizational members rooted in the communication of the evaluation system's function and purpose, which then has an impact on the implementation and results of the system.

The sociological variables of the context and interactions of all the members related to the organization build and develop the culture of the educational organization. The culture of educational organizations, as defined from an interpretive-ethnographic position, develops from internalized norms that "interlock into social formations" (Beuving & de Vries, 2015, p. 32) which the members of the organization use to build patterns of behavior (Heath & Street, 2008). The context of individual organizations encompasses the characteristics of the community, including (but not limited to) the socio-economic levels, the political power structures, and the ethnic-racial diversity, that would have an impact on the behaviors of the members within the

organization. The sociological variables of “status, power, sanction, and security” shape these interactions and the attitudes within and beyond the organization specifically around teacher evaluation, and “[t]o ignore the sociology of teacher evaluation is to fail to understand how systems actually work” (Peterson, 2004, p. 74). The impact on how the members of the organization will enact and/or react to the evaluation system depends on the intricate balance of the sociological nature of appraising teaching performance with the established culture related to the evaluation practices used by the organization (Cooper, Ehrensals, & Bromme, 2005; Nevo, 2006). The process of judging teacher performance related to identifying and supporting individual improvement of practice is a process that has impact on the academic-learning context of the organization and is based on the interpersonal relationships of the members in the organization (Darling-Hammond, 1990; Ingvarson & Rowe, 2008; Lavigne, 2014; Natriello, 1990; Nevo, 2006; Peterson, 2004).

Teachers and other related educational leaders must be cognizant of these complexities and variables that influence the purpose and function of evaluations so that evaluation systems work towards making an impact on teachers’ practice for the benefit of the whole organization (Nevo, 2006). Attention to culture and context also provides a sense of fairness and equity across the organization, which would support the system’s validity and credibility for assessing personnel for accountability purposes. When evaluators and personnel collaborate in productive ways to align the system with the goals and priorities foundational to the culture and context of the organization, the organization can build capacity and professionalism throughout the corps of employees that will drive the efficiency and productivity of the whole organization (Hargreaves & Fullan, 2012).

The body of literature on evaluation design and function includes the debate on whether an evaluation system can function for the dual purposes of formative and summative assessment of personnel performance within the same system (Cooper, et al., 2005; Darling-Hammond, 2013, 2014; DiPaola & Hoy, 2014; Huber & Skedsmo, 2016; Marzano, 2012; Popham, 2013). The research for both sides of this debate shows a wide range of support. The side of the debate that supports a distinction between the two purposes notes the need to include separate procedures and administrators because the process will work “best when the distinctions are clear to teachers” (Peterson, 2004, p. 68) in the organization. This differentiation of the two purposes depends on the way organizations conceptualize how to use the information from ratings, balance the contexts based on the sociological variables of the organization, and make judgments on how to use the information to determine the merit, worth, and value of the teachers’ practice (Darling-Hammond, 1990; Lavigne, 2014; Marzano, et al., 2011; McLaughlin, 1990).

The formative portion of the process addresses how teachers and organizational leaders measure the efficacy of personnel with tools and procedures that identify strengths and weaknesses of the teaching and learning that has an impact on productivity (i.e. student learning). The tools used to measure and address efficacy issues also provide the summative data necessary to address school reform requirements or mandates (Marzano, 2012). Evaluation systems combining these dual purposes would address the needs of the teacher (i.e. developing effective teaching practices) while also addressing the needs of the organization (i.e. raising student achievement) from a balanced sociological stance (Nevo, 2006; Peterson, 2004). In theory, systems that effectively address the dual purposes may have a formative impact on teachers and student performance while also serving the human resource purpose, but there are

few, if any, empirical studies that confirm that impact (MET Project, 2014; Myung & Martinez, 2013).

Dimensions of Evaluation Feedback Conferences

The research and literature on clinical supervision and evaluation provides the dimensions of effective evaluation-feedback conferences (Helm & St. Maurice, 2006; Hill & Grossman, 2013; Marshall, 2013). In essence, the evaluation feedback conference is an opportunity for an evaluator and teacher to unpack practice and student learning by having a two-way conversation about “what has been communicated throughout the evaluation period” so there “should be no surprises in the summary evaluation conference” (Helm & St. Maurice, 2006, p. 6). This conversation brings the teacher’s perspective on practice and the evaluator’s observation of practice to the center of the feedback conference to realize the greatest potential to change what happens in the teachers’ day-to-day instructional context (Arneson, 2015; Danielson, 2015; Hill & Grossman, 2013; Helm & St. Maurice, 2006; Kise, 2014; MET Project, 2015; Reilly, 2015; Renfro, 2014; Roussin & Zimmerman, 2014; Tuytens & Devos, 2012, 2014). As one component of the over-all evaluation process, the feedback conference functions on multiple levels based on observation evidence, stated purpose of the conference, and specific approach evaluators use to conduct the conference. Combined, the literature and research imply the evaluation process should include all these dimensions to have an impact on how teachers and evaluators perceive the effectiveness of evaluation process as a whole.

The literature and research on educational leadership, clinical supervision and evaluation provide considerations related to the various dimensions involved with conducting feedback conferences, for various combinations of functions and purposes. The following table is a

synthesis of this literature, showing the complexity of incorporating multiple dimensions, simultaneously, when conducting evaluation-observation feedback conferences:

Table 1

The topics evaluators focus on during feedback conversations:	Identified <u>purpose</u> for how evaluation data are used:		<u>Interpersonal</u> factors that impact feedback:	
		Summative		Formative
	Learning Outputs • focus on student learning outcomes	• To determine or assess personnel’s productivity, competency and accountability		• To develop personnel’s skills, knowledge and professional efficacy
	Learning Processes • focus on engagement and differentiated instructional practices	• for student assessment scores in relation to achievement levels for meeting learning standards and showing growth/meeting grade-level expectations		• for how the teacher develops student skills that support reaching benchmark learning goals and/or addresses gaps in learning
	Teaching Practice • focus on methodological and pedagogical decisions	• for responses to student work that reflects meeting the needs of all learners to reach expectations as set by benchmark or culminating assessments		• for student engagement strategies and instructional decisions that impact student learning; teachers’ response to student learning needs and goals
Professional Responsibility • focus on actions and decisions within the context and culture of teaching position	• for specific instructional practices and evidence of impact on student learning outcomes or achievement scores	• for developing teachers’ self-awareness and reflection on decisions and how those decisions impact student achievement/learning		
	• for evaluation ratings (i.e. scores) according to rubrics of teaching practice for meeting NYS Teaching Standards	• for using observation evidence, aligned with dimensions of teaching frameworks and standards, that identifies areas of development, needs/growth, and/or strengths • for developing professional relationships with colleagues/peers	<ul style="list-style-type: none"> • How the evaluator shows an understanding of content and instructional challenges particular to evaluation context • How the evaluator individualizes the evaluation content/procedures that promotes engagement in and accountability for the evaluation process as a whole • How the evaluator cultivates a culture of respect, recognition, appreciation and trust based on relationships between all stakeholders • How the evaluator balances the focus on school/district/state evaluation policies and/or requirements with maintenance of established and accepted goals/vision for student success • How the evaluator establishes credibility and validity of evaluations through organization, attention, and timeliness of accurate, equitable, and actionable appraisals of practice 	

Note. Synthesis based on Behrstock-Sherratt, et al., 2013; Cherasaro, Brodersen, Yanoski, Welp, & Reale, 2015; Costa & Garmston, 2002; Danielson, 2009, 2015; Danielson & McGreal, 2000; Darling-Hammond, 1990, 2013, 2014; DiPaola & Hoy, 2014; Donaldson & Donaldson, 2012; Frontier &

Mielke, 2016; Hall, 2019; Helm & St. Maurice, 2006; Hopkins, 2016; Ilgen, et al., 1979; JCSEE, 2009; Killion, 2015; LeFevre & Robinson, 2015; Lipton, et al., 2003; Marshall, 2013; Marzano, 2017; Marzano & Toth, 2013; Marzano, et al., 2011; MET Project, 2014, 2015; Myung & Martinez, 2013; Papay, 2012; Popham, 2013; Rigby, 2015; Scheeler, Ruhl, & McAfee, 2004; Stone & Heen, 2014; Stronge, 2007; Tuytens & Devos, 2012, 2014, 2017; Wiggins, 2012; Willis & Ingle, 2015)

The above table synthesizes the dimensions (i.e. topics discussed, purposes for collecting/giving feedback on observation, interpersonal behaviors and approaches to feedback/conferences) that should be foundational for feedback conferences to serve both summative and formative purposes while simultaneously functioning in evaluative and supervisory capacities cited in the credited studies and authors. Essentially, the table combines what the authors and studies propose as best practices for conducting observation feedback conferences; the over-arching focus, however, is on how *administrators-evaluators* should attend to multiple topics that simultaneously address dual purposes without attention to how teachers feel about those topics, purposes and behaviors/approaches.

The collected literature on this portion of the evaluation process (as noted in the literature synthesis table) is important to this study since an integral dimension of the evaluation feedback conference noted in the research questions focus on teachers' perceptions for how evaluators are adept at establishing a collaborative, constructive two-way dialogue that fulfills dual functions within the over-all evaluation system. As noted in the literature on evaluation system design (JCSEE, 2009; Scriven, 1991, 1994), feedback conferences can address dual purposes using evaluation data, but how the data are used for those purposes depend on the way evaluators conduct the conference. The interpersonal approaches, as a separate dimension that is more subjective in nature than what topics/purposes teachers objectively report on, highlight the variations in evaluators' stance, as supervisory or evaluative, which then impacts the receptiveness of the feedback for either purpose. The complexity of how to integrate all the

objective dimensions of topics and purpose, with consideration for how evaluators' interpersonal behaviors and approaches impact the conference itself, makes implementing all the dimensions effectively very difficult. What the literature and studies seem to provide are too many directions for evaluators, or teachers for that matter, to apply to the conference experience without clear distinctions about how each/all combinations will impact the effectiveness of the conference activity.

Over-all, this body of literature advocates for the feedback conference to provide time and opportunity for deep, pedagogical discussions about student and teacher learning for the intent of having an impact on a teacher's practice that in turn has an impact on students (Killion, 2015; Myung & Martinez, 2013; Popham, 2013; Tuytens & Devos, 2014). The feedback conference is also an opportunity for evaluators to bring issues or concerns to the attention of the teachers being evaluated that may have an impact on employment decisions (Steinberg & Donaldson, 2016; Weems & Rogers, 2010). The studies and research literature agree that giving and receiving feedback to be the cornerstone for effective performance evaluation conferences (Arneson, 2015; Behrstock-Sherratt, et al., 2013; Hall, 2019), and a large portion of studies provide suggestions for how evaluators or teachers can shift their existing paradigm for more effective and collaborative performance feedback conferences.

Many of the authors give supporting ideas, from the evaluator's position, on how to create an atmosphere of collaboration (Arneson, 2015; Brookhart & Moss, 2015; Combs, Harris, & Edmonson, 2015; Costa & Garmston, 2002; Danielson, 2015; Drago-Severson & Blum-DeStefano, 2014; Johnson, 2015; Kise, 2014; Lowenhaupt, 2014; MET Project, 2014; Myung & Martinez, 2013; Reilly, 2015) that will foster a two-way conversation. The crux of employing the strategies they promote is the ability of the evaluator/supervisor to develop the conversational

skillset that results in both participants feeling the conversation was effective for having an impact on the teachers' practice (Behrstock-Sherratt, et al., 2013; Benjamin, Yeager, & Simon, 2012; Stone & Heen, 2014). A number of studies focus on supporting administrators who are the lead evaluator or supervisor by coaching with question stems and prompts to use in particular situations (Lipton & Wellman, 2013; MET Project, 2014, 2015; Roussin & Zimmerman, 2014; Stone & Heen, 2014). These strategies give a framework as a means to develop evaluators' or supervisors' skills that would foster a collaborative relationship between them and teachers necessary for effective conferences. The over-all idea common across the literature is that effective feedback conferences relies on the ability of the evaluator or supervisor to cultivate a collaborative relationship that can work for both the summative and formative purposes of the over-all evaluation process. In theory (Ilgen, et al., 1979; Jawahar, 2010; Kinicki, et al., 2004; Kluger & DeNisi, 1996; Scheeler, et al., 2004), evaluators should develop the interpersonal skill-set that would help facilitate and support the kind of conversation that promotes the development of teacher motivation for taking accountability for student achievement, reciprocal trust between the evaluator and teacher about performance appraisal, and teachers' reflective abilities to assess practice to make changes, if necessary.

Another dimension of the interpersonal conference relationship is how evaluators engage teachers in thinking about observed practice. The literature and studies promote developing leadership skill-sets for providing feedback that would facilitate and support the recommendations for change which teachers can enact as a result of the conversation (Benjamin, et al., 2012; Danielson, 2015; Drago-Severson & Blum-DeStefano, 2014; Kise, 2014; Le Fevre & Robinson, 2015; Lowenhaupt, 2014; MET Project, 2015; Myung & Martinez, 2013; Rigby, 2015; Roussin & Zimmerman, 2014; Stone & Heen, 2014; Tuytens & Devos, 2014; Willis &

Ingle, 2015). The suggestions focus on educational leadership practices based on developing trust with teachers about decision making, engaging teachers in collaborative thinking about practice, and listening to teachers that shows reflection on current practices. The studies even go as far as providing set protocols and suggestions for evaluators who are unsure of how to approach feedback conferences that would support and facilitate teachers' understanding and acceptance of responsibility for instructional practices (Benjamin et al., 2012; Combs, et al., 2015; Danielson, 2009, 2015; Darling-Hammond, 2014; Myung & Martinez, 2013; Peng & Lin, 2016; Roussin & Zimmerman, 2014; Stone & Heen, 2014).

There are also specific studies that give suggestions on how to construct this two-way conversation that would help develop evaluators' feedback-giving abilities and teachers' receptivity of the feedback. Myung & Martinez's (2013) brief on enhancing the feedback experience, the MET Project (2014, 2015) on evaluation feedback training and Stone & Heen's (2014) text on giving and receiving feedback focus on how evaluators should use conversation protocols, and they endorse having evaluators collaboratively practice using the protocols with other evaluators and teachers to ensure validity of the ratings and feedback. The primary focus of these studies is on developing the evaluators' skills, but even if evaluators follow how to co-construct the conversation in the suggested ways, the observation conversation presumes teachers do not already have the skills, or have shown to lack the skills, of reflection in-on-of effective practice (Drago-Severson & Blum-DeStefano, 2014; Schön, 1987; Tuytens & Devos, 2014). Over-all, the studies and literature support the need for evaluators to provide actionable feedback and outline protocols for effectively presenting that feedback to prompt change in practice, but the impact on teachers' practice that directly results from this feedback conversation needs more empirical study. Some of the studies provide protocols and suggestions for

evaluators who are unsure of how to approach feedback conferences that would support and facilitate teachers' trust, understanding and acceptance of responsibility for instructional practices (Benjamin, et al., 2012; Combs, et al., 2015; Costa & Garmston, 2002; Danielson, 2009, 2015; Darling-Hammond, 2014; Myung & Martinez, 2013; Peng & Lin, 2016; Roussin & Zimmerman, 2014; Stone & Heen, 2014). However, for the evaluators who follow a pre-designed structure or script to develop inter-personal skills necessary for building trust with colleagues is not a guarantee that the end result will be a trusting relationship that can impact teachers' practice (Behrstock-Sherratt, et al., 2013; Costa & Garmston, 2002; MET Project, 2014; Myung & Martinez, 2013).

In general, there are few studies on teachers' perceptions of evaluators' skills and/or approaches when conducting the evaluation feedback conference that would cultivate the level of trust that results in the feedback having an impact on the teachers' practice. These studies (Finster & Milanowski, 2018; Hopkins, 2016; Range, Young & Hvidston, 2013; Tuytens & Devos, 2012, 2017; Zimmerman & Deckert-Pelton, 2003) surveyed and interviewed teachers to explore their perceptions primarily about the administrators'/evaluators' *roles* in the evaluation systems in place, not specifically on the approaches evaluators use to conduct the feedback conference. The findings from the studies over-all coalesce around themes for how teachers respond to evaluators who show credibility, fairness and trustworthiness; however, these studies also note teachers reported feeling distrust toward the evaluator which impacts the evaluation process when there is an absence of these qualities with an evaluator. These conclusions drawn from the data findings highlight the impact mistrust or lack of credibility has on teachers' perceptions of conference usefulness, which imply tensions exist between teachers and evaluators in the evaluation context that interfere with the utility of the evaluation. Each study

proposes the need for further research that focuses on *teachers'* perceptions of evaluators' approaches and the nature of the impact those approaches have on the effectiveness of evaluation conferences as a way to address tensions, build collaborative relationships and develop trust in evaluation systems over-all.

Appraisal, judgement and understanding of teaching practices, apart from student assessment scores, are the basis for teacher evaluation systems, and how educational organizations implement procedures for the appraisal, judgement and understanding of teaching contribute convey the culture and norms to members of the educational organization. (Danielson & McGreal, 2000; Darling-Hammond, 2013; DiPaola & Hoy, 2014; Marzano, 2017; Marzano, et al., 2011; Popham, 2013; Scriven, 1991; Stronge, 2007). Usually, there is a general agreement across school contexts and cultures on the requisite duties, responsibilities, and skill-sets of teachers and administrators that contribute to those norms based on a set of standards; the standards are used as the criteria for judging the content and competency of the work that teachers do and are aligned with over-all school goals and priorities. Credibility for a system that uses professional teaching standards comes from how the standards synthesize multiple sources and studies on effective teaching practices and teacher dispositions to determine what constitutes quality teaching and teacher quality (Danielson & McGreal, 2000; Darling-Hammond, 2013; Ingvarson & Rowe, 2008; InTASC, 2011; Marzano, et al., 2011; NBPTS, 2002; NYSED, 2011; Popham, 2013; Stronge, 2007). Using teaching standards as an integral part of the system design, in theory, requires organizational leaders and teachers to agree on methods of evidence collection (i.e. observation rubrics, student work) that would show how teachers are meeting adopted teaching standards at proficient levels of competency (Danielson & McGreal, 2000; Darling-Hammond, 2013; DiPaola & Hoy, 2014; Marzano, et al., 2011; Popham, 2013; Stronge, 2007).

Through the use of standards and identified teaching competencies, organizations establish credibility for the evaluation system because it meets professional and institutional criteria for competency. From the theoretical standpoint, professional standards and performance criteria can be a gauge to “learn what there is to learn” (Scriven, 1991, p. 257) about how personnel function in relation to the goals, objectives and priorities of an organization. Teaching standards are also used to show an explicit connection between student performance, such as student work samples and assessment scores (the ‘product’ of teacher performance), to the content of what the teacher presents. Both evaluators and teachers need to understand the criteria, scope and vision of effective teaching practices and dispositions that are research-based since standards support and articulate concepts of what teachers should know, how teachers know, and how teachers are able to act upon that knowledge at proficient levels of performance in the classroom (Danielson & McGreal, 2000; Ingvarson & Rowe, 2008; InTASC, 2011; NBPTS, 2002; NYSED, 2011).

As part of an effective evaluation system design, organizations need to understand the performance indicators of adopted teaching standards and how to recognize the levels of proficiency of teaching practices aligned with the standards when observed (NBPTS, 2002). As a function of evaluation, teaching standards can be aligned with student learning standards through student standardized assessments; teachers’ performance can be judged competent or not based on the evidence from those student assessment scores for the impact of aligned instructional practices. As a function of supervision, teaching standards help evaluators observe teachers’ behavior for evidence of meeting the performance indicators of teaching competencies aligned with what research says are best practices. When there is evidence of not meeting the standards, evaluators and teachers use the data to understand what areas of practice would need

improvement or strengthening to have a greater impact on student performance. Evaluation theory and the research literature note that effective systems should align teaching standards with the evaluation processes in place since the content of teachers' practice can be evaluated equitably when standards are applied as the means for assessing, appraising and judging the competency of that practice.

There are presently two general teaching standards documents, the Council of Chief State School Officers' *InTASC Model Core Teaching Standards* (2011) and National Board for Professional Teaching Standards (2002), cited as guiding resources to develop programs and evaluation tools of effective/quality teaching (InTASC, 2011; NBPTS, 2002). Evaluation systems use the standards and performance indicators linked to the standards to determine the level of teaching proficiency in authentic teaching contexts, and the assessment of the impact of that proficiency on student learning becomes linked with the productivity of the organization (Danielson & McGreal, 2000; InTASC, 2011; Ingvarson & Rowe, 2008; NBPTS, 2002).

Standards, in the sense of measures, are tools we use constantly in making judgements in many areas of life and work, whether measuring length, evaluating writing, critiquing restaurants or measuring professional performance. Standards provide the necessary context of shared meanings and values for fair, reliable and useful judgements to be made (Ingvarson & Rowe, 2008, p. 16).

Used as a developmental resource, teaching standards provide a basis for evaluation frameworks that use pedagogical language and conceptual understanding of effective practices that cut across content and grade levels (InTASC, 2011; NBPTS, 2002). With teaching standards in place, organizations can develop evaluation policies that will "honor the complexities and demands of teaching" and "focus on teacher work and the difficult issues that accomplished teachers confront on a regular basis" (NBPTS, 2002, p. 1).

Teaching standards also serve the purpose of categorizing teaching behaviors and dispositions, which educational organizations use to identify areas of concern or success, and provide a means to develop goals and objectives related to the organization's purpose and instructional needs (Popham, 2013; Scriven, 1995; Tuytens & Devos, 2014). Through this structural lens, teaching standards support organizations' evaluation processes for making judgments on the merit, worth and value of teachers' practice against an accepted paradigm of teacher quality and quality teaching (Ingvarson & Rowe, 2008; Marchant, et al., 2015). An effective evaluation system uses teaching standards to ensure equity, consistency and balance when making human resource decisions (the summative purpose) and developing professional support systems (the formative purpose).

Frameworks or models of quality teaching use the teaching standards to articulate areas of common performance domains and indicators used in development of evaluation tools (Danielson & McGreal, 2000; DiPaola & Hoy, 2014; Marzano, et al., 2011; Muñoz, et al., 2013; Stronge, 2007). The domains and indicators of quality teaching rely on the research consensus regarding how teachers comprehend the content of their specific discipline, use an array of instructional strategies to fit specific teaching contexts, adapt instruction for student needs, assess student and own understanding, and reflect on practice for student academic growth (Muñoz, et al., 2013; Schön, 1983, 1987; Scriven, 1991).

Feedback Impact on Teachers' Practice

Broadly, the term *feedback* applies to learning about self or a situation through a process or procedure of examining evidence from performance data related to the context of the professional situation. For many of the contexts where performance evaluation is necessary, feedback applies to how various systems or organizations effectively give and receive it, and

how feedback specifically has an impact on the practice of an individual (Benjamin, et al., 2012; Hill & Grossman, 2013; Ilgen, et al., 1979; Jawahar, 2010; Kinicki, et al., 2004; Kluger & DeNisi, 1996; Peng & Lin, 2016; Stone & Heen, 2014; Wiggins, 2012; Zingoni, 2017). Through feedback, as an intervention that makes an impact on performance, organizational leaders and employees collaborate to find ways to implement change in practices, develop culture or climate of an organization, or reform already established organizations (Hill & Grossman, 2013; Ilgen, et al., 1979; Kinicki, et al., 2004; Kluger & DeNisi, 1996, 1998; Zingoni, 2017).

Within this general field of study on feedback, there is a small but growing body of literature on how performance-evaluation feedback has an impact on professional teaching practices (Cherasaro, et al., 2016; Costa & Garmston, 2002; Danielson, 2009; Hall, 2019; Hill & Grossman, 2013; Hopkins, 2016; Killion, 2015; Kimball, 2003; Khachatryan, 2015; Scheeler, et al., 2004; Scriven, 1995; Quintelier, Vanhoof, & De Maeyer, 2018). These studies note the necessity for evaluators to give feedback that should provide actionable suggestions or reflections on a teacher's practice (Brookhart & Moss, 2015; Killion, 2015; MET Project, 2012, 2014). Feedback during the post-observation conference, if given in a collaborative and constructive way as noted in the literature, can lead to improvements in teachers' practice that will in turn improve student achievement (Roussin & Zimmerman, 2014), which meets the standards of evaluation utility and accuracy (JCSEE, 2009).

Much of the supervision and evaluation literature focuses on how the *administrator or evaluator* can be the dynamic actor who influences the teacher to accept, reflect and use the information from performance appraisals. This dimension relies on the development of leadership skill-sets for effectively implementing conference procedures that would facilitate and support recommendations for change which teachers can enact as a result of the feedback

conversation (Costa & Garmston, 2002; Danielson, 2015; Drago-Severson & Blum-DeStefano, 2014; Kise, 2014; Le Fevre & Robinson, 2015; Lowenhaupt, 2014; MET Project, 2015; Myung & Martinez, 2013; Rigby, 2015; Roussin & Zimmerman, 2014; Tuytens & Devos, 2014; Willis & Ingle, 2015). The implied concept is that the feedback conferences can be considered effective if evaluators conduct feedback conferences in such a way for utility and accuracy (JCSEE, 2009) that prompt or motivate teachers to take ownership of the evaluation process and accept accountability for student learning (Calabrese, et al., 2004; Costa & Garmston, 2002; Danielson & McGreal, 2000; Darling-Hammond, 2013, 2014; Helm & St. Maurice, 2006; Scriven, 1995). However, these studies note the difficulty of qualifying the various ways *teachers* receive, engage and accept feedback about their performance in ways that would show accountability for their instructional practice, engagement in professional development, and ownership of the over-all evaluation process (Khachatryan, 2015; Tuytens & Devos, 2012, 2014, 2017).

The implied end result of the feedback conference is for teachers to accept and make changes to practice that come from the evaluators' observation and suggestions, especially when given the *opportunity* to articulate and reflect on this process of change during the post-observation conversation. Many studies note that creating an *opportunity* in the post-observation conversation for teachers to articulate and reflect on this process of change can lead to improvements in practice (Arneson, 2015; Brookhart & Moss, 2015; Costa & Garmston, 2002; Lortie, 1975; Marzano, 2012; Reilly, 2015; Roussin & Zimmerman, 2014; Schön, 1983, 1987; Tuytens & Devos, 2014; Zingoni, 2017). However, the studies also note that negative perceptions of initiating change in practice, held by teachers and some administrators/evaluators, still exists (Arneson, 2015; Brookhart & Moss, 2015; Costa & Garmston, 2002; Marzano, 2012;

Reilly, 2015; Roussin & Zimmerman, 2014; Tuytens & Devos, 2014). The impact feedback has on the way teachers self-assess and reflect on their practice becomes part of the means to identify how evaluations are able to serve summative and formative purposes as required by the APPR law. What teachers feel about and do with feedback on their teaching becomes the outcome of the evaluation process.

Summary

The body of research literature on this topic puts forth integral conditions that organizations should consider for the development and implementation of evaluation systems to ensure evaluation feedback conferences are effective for teachers and the organization. However, this existing body of literature predominately focuses on how *educational leaders and supervisors*, acting in the capacity as evaluators of teachers' performance, should conduct conversations during conferences about observation data that includes how data contributes to an over-all evaluation rating. The literature also predominately addresses how *educational leaders-as-evaluators* navigate evaluation-feedback conferences for both summative and formative purposes. What is notable in the existing literature is the paucity of research that focuses on how *teachers* perceive what makes conferences effective and how they make sense of the way evaluators conduct evaluation conferences.

Specifically, this study will look to the dimensions for strategies evaluators use during the feedback conferences noted in the body of literature. The small body of existing research literature exploring teachers' affective responses to evaluation conferences and evaluators' approaches to conducting those conferences uses feedback theory as a possible lens to explain teachers' reactions to over-all performance feedback; however, those studies do not specifically focus on how evaluators present the feedback during evaluation conferences and whether their

approaches have an impact on teachers' practice. The literature also does not address teachers' affective responses to evaluation of their practice when feedback is meant to serve dual purposes. This study addresses this gap in the research literature by focusing on teachers' experiences and perceptions of feedback as part of the evaluation process, and endeavors to include teachers' voices on the topic of what makes evaluation feedback conferences effective for both teachers and administrators.

Chapter 3

Methods

Overview of Research Design, Instrumentation, Implementation, Data Collection

Early in my research apprenticeship study (RAP), it became evident that exploring teachers' perspectives on evaluation conferences would be a complicated endeavor if just one research method was used, which would also impact the validity and consistency of the findings. The RAP interviews provided insights on multi-faceted ways teachers frame their understanding of evaluation conferences, complicated by layers of personal (i.e. identity as an effective teacher), professional (i.e. dealing with multiple iterations of performance reviews) and contextual (i.e. how districts implement APPRs and interpersonal relationships/ dynamics within districts) experiences; however, conducting more qualitative interviews to continue my research would still limit the sample of participants required to further this research. Once I developed the dissertation research questions, I made the purposeful decision to take a mixed methods approach to explore the evaluation conference phenomenon. By making such a decision, I would be opening the study to include qualitative and quantitative approaches for collecting data, thereby capturing multiple perspectives from varied positions (subjective and objective) on a complex phenomenon, while also increasing the generalizability and consistency of the over-all findings (Luyt, 2012; Onwuegbuzie, Bustamante, & Nelson, 2010).

Research Focus and Questions

This study explored how evaluators conduct effective evaluation feedback conferences from the teachers' perspectives, and how/which specific approaches teachers perceive as having utility for improving teachers' practice and increasing their ownership of the evaluation process. The findings from exploring the nature of evaluators' approaches and teachers' perceptions of

those approaches may provide insight on the tension teachers and evaluators experience related to the APPR evaluation conferences over-all. Understanding the dynamics involved with evaluation feedback conferences between teachers and evaluators has the potential to move the field toward developing a theory related to how an evaluator's approach to conducting evaluation feedback conferences contributes to the effectiveness of the over-all evaluation process.

The initial research apprenticeship study asked teachers about their APPR evaluation and post-observation feedback conference experiences with evaluators responsible for observing, rating, and supporting them. Two initial, over-arching questions attempted to capture these experiences, which were used to develop the interview guide (see Appendix A: RAP Interview Guide):

1. What are the over-all experiences teachers are having with APPR evaluation conferences that have an impact on their practice?
2. Does the process of and approach to the APPR evaluation conferences facilitate teachers taking ownership/agency of their practice and APPR evaluation process?

The interview study design opened up the discussion between the researcher and participants on the topics of evaluation conferences, with the assumptions that participants would organically respond with narratives that would explicitly reflect what they felt were effective or ineffective approaches evaluators take in conferences. It was assumed that participants would address the formative-summative tension that the NYS APPR Law implies. Once analyzed, the interview data revealed areas for further research on how teachers respond to professional contextual situations and specific approaches evaluators use when conducting evaluation conferences that impact the conferences' effectiveness, and not necessarily on the formative-summative issue.

The two initial research questions thus provided the basis for the next phase of research. The survey and q-sort components were developed to address similar research questions:

1. What are teachers' experiences with how evaluators conduct evaluation feedback conferences?
2. Is there a connection between the way evaluators conduct feedback conferences, as reported by teachers, and how those conferences affect teachers' practice?
3. Based on teachers' reports, under what circumstances (if any) is it possible for evaluation feedback conferences to serve both summative and formative purposes?

When developing the research questions to guide the survey, the first apprenticeship study question was divided between exploring teachers' experiences further by focusing on reporting of specific actions by evaluators and reporting on how that experience impacts their practice. The latter half of the first apprenticeship question became combined with the second apprenticeship question to focus on the impact teachers report the approaches may/may not have on their practice. I expected that the second question of the dissertation study would elicit more explicit and objective responses that would provide perspectives on the context and culture of teachers' experiences that make feedback conferences effective for teachers and evaluators.

The third dissertation research question addresses an issue implied in the research literature on evaluation and supervision and the current APPR law: an evaluation process that attempts to serve multiple functions and purposes simultaneously. I expected that the feedback conference might be the key to addressing the tensions participants feel in the evaluation process when the process serves multiple functions (supervision and evaluation) and purposes (formative and summative). In all, the set of initial research questions and the dissertation study research questions reflect the complexity of teachers' experiences with the evaluation feedback conference phenomenon. To capture this complexity of teachers' perspectives and sense-making around the phenomenon, I felt mixed methods would suit this study best because mixed

approaches would allow the ‘truth’ of teachers’ experiences to be represented by the data collected, showing multiple facets of their perspectives without distorting that ‘truth.’

Research Study Design

My dissertation research study explicitly combined qualitative, quantitative, and mixed approaches to data collection for an over-all pragmatic mixed methods design. A prominent feature of pragmatic mixed methods is how each approach gathers data in methodologically different ways, reflecting an iterative process to develop/inform the collecting of data with the other approaches; a mixed methods research study that takes this pragmatic/practical and cyclical approach to using qualitative and quantitative methods can “provide different, but complementary, data” (Luyt, 2012, p. 296). Following a QUAL-QUANT-MIXED design-framework (Onwuegbuzie & Johnson, 2006), I used Naturalistic Inquiry semi-structured interviews as the qualitative approach, Survey Research as the quantitative approach, and Q-Methodology Q-Sorts as the mixed/hybrid approach. These approaches fit together as an exploratory-sequential-conversion model by converting narrative interview responses, combined with themes from evaluation-supervision literature, into objective survey items for scaled responses, and then reframing the survey findings as subjective q-sort statements to be rank-ordered from individual perspectives (Creswell, 2014; Leech & Onwuegbuzie, 2009; McKeown & Thomas, 2014; Onwuegbuzie & Johnson, 2006; Teddlie & Tashakkori, 2006).

Each distinct approach collected separate sets of data rooted in the paradigm in which it ‘belongs’ (i.e. qualitative paradigm, quantitative paradigm and mixed methods paradigm), but from epistemologically different stances or positions. The qualitative approach used semi-structured interviews to collect narrative data and a constructivist position for interpreting and analyzing subjective discourse inherent in participants’ narratives. The quantitative approach

used a multi-sectioned/item survey to collect objectively scaled data and an empirical/statistical position for interpreting and analyzing factors and relationships among factors based on participants' objective reporting of experiences. The Q-Method/mixed approach used ranked/distributed statements to collect subjective data and constructed responses, and relied on both qualitative/constructivist and quantitative/statistical positions for identifying, analyzing and interpreting participants' ranking of statements. When used as separate lenses to analyze and interpret participants' experiences, the data contributed by each approach provided a focused understanding of the research topic because of the paradigm's methodology/framework for using the distinct approach. When the three approaches were taken together and used as a collective lens to analyze and interpret the data, those seemingly disparate sets of data provided a fuller picture that mono-methods/approaches were not able to accomplish (Eden, Donaldson, & Walker, 2005; Ellingsen, Størksen, & Stevens, 2010; McKeown & Thomas, 2014; Onwuegbuzie & Leech, 2005; Ramlo, 2016; Watts & Stenner, 2005). The following table provides an overview of each method used:

Table 2: QUAL-QUANT-MIXED Paradigms

<u>Qualitative</u> <i>Naturalistic Inquiry</i>	<u>Quantitative</u> <i>R Research Method</i>	<u>Hybrid</u> <i>Q Research Method</i>
<u>Research Goal</u> To collect participants' <i>subjective narratives</i> of <u>EXPERIENCES</u> with evaluators and impact of conferences/ evaluations on practice	<u>Research Goal</u> To collect participants' <i>objective reporting</i> of evaluation conference <u>EXPERIENCES</u> and behaviors of evaluators from own direct experiences	<u>Research Goal</u> To collect participants' <i>subjective ranking</i> of items most associated with individual <u>EXPERIENCES</u> for how evaluation conferences are conducted
<u>Research Output</u> Data on affective- <i>subjective</i> themes related only to participants' experiences	<u>Research Output</u> Data on <i>objective</i> factors and variables related to conference effectiveness measures reported by participants	<u>Research Output</u> Data on <i>objectively factored subjective perspectives</i> of participants on effective conference approaches
<u>Overall contributions:</u> data provided features of	<u>Overall contributions:</u> data provided objective behaviors	<u>Overall contributions:</u> data provided which composite

<p>evaluation feedback conferences that participants subjectively identified in response to open-ended prompts as most prominent and important in evaluators' conduct of feedback conferences.</p>	<p>respondents reported most accurately described how evaluators conducted their most recent evaluation conference; what composite factors most accurately and reliably represented these behaviors; and how strongly or weakly these factors correlated with participants' overall assessments of the conference's effectiveness.</p>	<p>descriptions of evaluators' behaviors (drawn from earlier data analyses) participants subjectively identified as most positively, negatively or weakly associated with effective conferences; whether different groups of teachers are more sensitive to different factors.</p>
---	--	---

The research design sequence began with interview data collected from my research apprenticeship project exploring how evaluation feedback conferences have an impact on teachers' practice (qualitative approach). I conducted eight semi-structured interviews, as a 'purely' qualitative approach, which provided enough narrative text for strong themes to emerge (Saldaña, 2013) related to how teachers *subjectively felt* when evaluated and during/after evaluation feedback conferences. The narrative data analysis showed teachers articulating complex social, political and contextual perspectives regarding experiences with evaluation conferences over-all. As strong as the themes seemed to be, the interview data/fieldnotes did not generate the amount of data that could be used to make broader assertions about effectiveness of the evaluation feedback specifically. After analysis and interpretation of the RAP interview data, I decided to further extend my research on teachers' perceptions of the evaluation feedback conferences with a broader sample group through survey research.

For the quantitative portion of my dissertation study, I quantitized the qualitative interview data, research literature, a previous iteration of a Teacher Evaluation Survey (see Appendix B: Syracuse University Study Council Teacher Evaluation Survey, 1995) and drafts of a simulation rubric to construct the survey instrument (Teddlie & Tashakkori, 2006). The Evaluation Feedback Conference Survey, as a 'purely' quantitative approach, used the themes

from the interviews and other sources of information to construct the sections of the survey (prompts and items) which asked participants to *objectively report* on conference topics, general and specific approaches evaluators take in conferences, school/district culture and context variables related to conferences, and extent of effectiveness for the feedback conference. Out of 98 invited districts, five agreed to participate; out of the five districts, one district contributed zero responses and the other four districts contributed 58 full or partial responses. Over-all, a total of 39 usable full responses were entered into SPSS software for statistical factor analysis. The relatively small data set showed statistically significant correlations and reliability between and within factors, but the low number of participants for the survey still cast the validity of the factors into question for having generalizability across the population for the survey.

As part of the research study design, I initially planned on using a focus group to further analyze the survey data, and I employed a Q Methodological approach to collect the responses from the focus group. As a third stage of the study, the Focus Group Q-Sorts used a ‘hybrid’ of qualitative and quantitative instrumentation, data collection and analysis (McKeown & Thomas, 2014; Stephenson, 2014) which provided a means to statistically analyze the subjective perspectives of an even smaller number of participants on the same themes and factors that emerged from the interviews and survey. The approach relied on individual participants in the focus group ranking subjective statements on the shared phenomenon of evaluation conference experiences. Participants also provided their rationalizations for ranking statements they felt to be ‘most’ to ‘least’ aligned with their individual perspectives of what contributed to, detracted from or has minimal impact on a conference’s effectiveness; the data collected from this ranking activity are called ‘q-sorts.’

The collected q-sorts were statistically analyzed using PQMethod software (Schmolck, 2014) which performs a factor analysis on the ranked statements by correlating the individual q-sorts for factors that represent shared perspectives on the q-sort topic. The q-sorting process also included collecting qualitative-constructed text from the participants on their individual q-sorts; these data were coded and added after statistical analysis to the representative factors for interpretation purposes. This third approach brings together the subjective way participants perceive the effectiveness of evaluators' approaches with the empirical, objective statistical analyses of that shared perspective. The collected data from the q-sorts provided clarification on the research questions while also complementing the already collected data from interviews (qualitative) and survey responses (quantitative). The following table provides an over-view of the implementation, sample populations and analysis-interpretation for each of the methods:

Table 3: Over-view of Research Approaches

<u>Qualitative</u> <i>Naturalistic Inquiry</i>	<u>Quantitative</u> <i>R Research Method</i>	<u>Mixed/Hybrid</u> <i>Q Research Method</i>
Semi-Structured Interviews: face-to-face meetings 40-65 minutes Sites--Participant's choice, after school hours	Survey: online (Qualtrics) 10-20 minutes Distributed via individual district's email	Q-Sorts: Small Focus Groups 90 minutes Site—neutral, after school hours
<u>Participants:</u> Convenience sample/word of mouth 8 teachers/all women Mix of tenured/untentured Mix of rural, suburban, urban	<u>Participants:</u> Random Sample (schools) 39 individual teacher responses, anonymous Mix of tenured/untentured Mix of rural, suburban, urban	<u>Participants:</u> Convenience sample/word of mouth 13 teachers (includes researcher)/mixed gender Mix of tenured/untentured Mix of rural, suburban, urban
<u>Data Analysis:</u> Qualitative coding Interview transcripts + fieldnotes= individual 'codebooks'	<u>Data Analysis:</u> Exploratory Factor Analysis, Correlations, Reliability SPSS reports including participant constructed responses	<u>Data Analysis:</u> Mix Factor Analysis/Qualitative coding PQMethod reports and participant constructed responses

The construction of the research instrumentation, process of data collection and initial steps of data analysis for each research approach follows:

Qualitative Paradigm: Naturalistic Inquiry—Semi-Structured Interviews

The decision to use semi-structured interviews for the first phase of the study was a result of document analysis and literature review for my research apprenticeship project on teachers' perspectives of evaluation feedback conferences. My over-all program coursework, including research methods, up to the time of conducting the apprenticeship project always related to what was being discussed and 'felt' personally and professionally (see Personal Subjectivity Statement at the end of this chapter) in the field. Because of the inherent subjectivity of the topic, a qualitative approach seemed to suit best as a means to 'enter' teachers' voices into the existing academic conversation (Fine, Weis, Weseen, & Wong, 2000; Kim, 2016; Patton, 2015; Rosenblatt, 2003; Seidman, 2006). In contrast to quantitative studies that remove individual context from consideration in the deductive analysis of data, naturalistic qualitative inquiry "highlights and deciphers context when interpreting findings" that "elevates context as critical to understanding" the nature of the phenomenon (Patton, 2015, p. 69). As the method to collect qualitative data, interviews have "increasingly democratized experiential information" (Gubrium & Holstein, 2000) that can lead the researcher to a more "conscious awareness of the power of the social and organizational context of people's experiences" (Seidman, 2006, p. 130).

With this awareness in mind, I crafted the interview questions during the RAP study design process to be open and flexible to allow co-construction of meaning to emerge as participants responded to questions and articulated perceptions of lived experiences (Rosenblatt, 2003; Seidman, 2006). Such a pragmatic approach provided a framework to determine the interview question guide (see Appendix A: RAP Interview Guide). As a qualitative approach,

pragmatism focuses the purpose of the interviews to elicit “[s]traightforward questions about real-world issues aimed at getting straightforward answers that can yield practical and useful insights” (Patton, 2015, Exhibit 7.3, pg. 436). The interview guide used a semi-structured format (Gubrium & Holstein, 1997; Kim, 2016; Patton, 2015; Seidman, 2006) for the purpose of asking teachers about ‘real-life’ experiences with feedback conferences and how those experiences had an impact on practice. The care and consideration when collecting narrative data relied on crafting interview questions that would open the inquiry for reflective examination by the participants and reflexive interpretation by the researcher (Rosenblatt, 2003; Seidman, 2006). I purposefully used the guide as the way to focus the interview for the best and most respectful use of participants’ time; I also used the guide to ensure the specific topic of evaluation feedback conferences was addressed within a limited time while also allowing time for participants to fully respond in ways that were meaningful for them.

After receiving Institutional Review Board (IRB) approval, I personally recruited eight teachers for individual, one-time semi-structured interviews. The participant group was a convenience sampling, found by word-of-mouth and by personally asking colleagues to participate as a professional courtesy. By using this method to populate my sample group, my interview participants have a limited range of personal demographics; all eight participants identify as white and female. However, I specifically asked this group to participate for the range of professional demographics; the participating teachers are employed across multiple districts (rural, suburban, and urban), from a range of grade levels (grades k-12) and across content/grade levels (special education at the elementary level, science, mathematics, and English). They have a range of teaching experience, from three years to 30 years of service. Since I knew all the participants on a professional level, I also knew they have a range of experiences with various

configurations of APPR evaluation systems, as well as with multiple evaluators within and between districts.

The interviews ranged from 40 minutes to 65 minutes, meeting in public places of the participants' choosing, and each interview was transcribed anonymously. I recorded each interview on a personal recording device (computer voice program and hand-held voice recorder). The audio files were transcribed anonymously by a non-research administrative assistant into Microsoft Word documents, and then I added fieldnotes to the transcriptions for coding purposes. Once transcribed and fieldnotes added, I organized the narrative data to create individual 'codebooks' for data analysis and interpretation purposes.

As the first cycle of coding (Saldaña, 2013), I purposefully assigned categories for how the narrative responses described the objective portion of what 'happened' in the conferences and other related tasks that participants felt impacted how the conference was conducted. I assigned the data to categories based on what participants articulated about their experiences related to individual districts' adherence to the APPR law and over-all evaluation conference experiences, not on a singular particular incident of an evaluation conference. All the participants' narratives presented evidence of variations in the district cultures and evaluation processes (context) which seem related to implementation of state-approved teaching practice rubrics as part of the district's APPR plan. I used the following categories to code data:

- Co-Construction of rubric meaning [through evidence collection/presentation, conversations and PD]
- Observation Feedback
- Paperwork and Procedure [that affects usefulness of evaluation]

In general, I understood the narratives as expressing perceptions of underlying external constructs for the influence school culture and context, political oversight of evaluation

processes (i.e. APPR legislation) and school/district/state policies that were out of their control as having some impact on how they engaged in the process. The participants did not necessarily share exact same experiences with these constructs, but in some way each participant did articulate instances of culture, context, policy, and politics playing roles in the effectiveness of the conference experience. An underlying narrative related to policy and politics, namely the way their evaluations seemed skewed by student growth scores and ratings from observations that only pertained to how teachers were addressing ‘test-prep,’ may have negatively shaped their over-all perceptions of the feedback conference experience and how they felt toward evaluators’ approaches to the feedback conference.

For the second pass in the first cycle of coding, I assigned categories to participants’ reflections that showed affective patterns related to how they internalized the impact of the conference experiences on their practice. As with the first cycle of codes, I categorized statements for explicitly subjective discourse on the portion of the conferences which participants noted as being more personal for them:

- Collegial Conversations
- Power-Trust-Control
- Collegial Understanding [including adjustment of observed practice for specific administrator/evaluator perspective/requirements]
- Authenticity + Appreciation
- Self-Efficacy/agency/worth of professional self [including how affected by administrator/evaluator assignments/changes in who is doing evaluation]
- Professional self =self-assessment of practice and self-report of effectiveness [including comparing self to others in other districts]

These subjective categories differed from the objective categories by how participants perceived the evaluation process, not just the conference, reflects a narrative about them as teachers. The categories focus on themes of interpersonal relationships, professional efficacy, and personal reflectivity/reflexivity that participants noted as what makes them identify as ‘teachers,’ but feel

that the evaluation process, most notably the feedback conferences, diminishes and does not account for what they feel they actually do in the classroom day-to-day.

Using affective coding methods (Saldaña, 2013), a second cycle of coding analyzed the individual codebooks for similarities and differences for how the participants articulated details of their experiences and expressed personal perspectives. During the second cycle of coding, I refined the objective and subjective categories to address the research questions and explore the hypotheses related to the questions of the RAP. I re-themed the objective reporting and subjective reflections for how the participants' responses could provide a framework for conducting feedback conferences for cultivating engagement in the process over-all. The second cycle of coding generated more applicable categories and themes towards addressing the research questions: the two research questions became the over-arching categories (bold-bulleted), with the subcategories (*) and themes (+) refining the categories and themes from the first cycle of coding:

- ***Conference experiences in general (responses for research question 1)***
 - *Evaluation methods-protocols teachers and evaluators follow
 - +consistencies/inconsistencies of evaluation methods within district/between districts
 - +use of protocols to gather evidence by/with evaluators
 - *Authenticity of evaluation systems
 - + forced authenticity of evaluation—teach to fit the rubric
 - + teachers' own means of self-evaluation
 - *District-State mandated approaches to conferences
 - + 'teaching to rubric' concept as authentic assessment
 - +consistency/inconsistency of rubric application within/between evaluators/districts
- ***Specific process/approaches that affected teachers' ownership/agency (responses to research question 2)***
 - *Control of the evaluation context
 - + trust of teachers with/for evaluators and the system of evaluation
 - + collaborative relationship with evaluators

- *Roles given/taken by individuals in evaluation conferences
 - + expectations of teacher/evaluator for conferences/system
 - + communication between teacher and evaluator about practice and evaluations
- *Perceptions of self, agency and responsibility of evaluation and practice
 - + ‘proving’ self to others to fulfill perceived obligations of system/evaluator
 - + respect/lack of respect for professional knowledge of teacher by evaluator/system

By coding the data with a research question-lens, the participants’ voices emerged as the dominant narrative about the impact evaluator’s approaches to the feedback conference have on impacting teachers’ practice.

Once interpreted and analyzed, the culminating step for the RAP was the research report. The resulting report included discussion of themes related to APPR requirements, culture, and context of the participants’ evaluation experiences; the report also provided the participants’ unscrubbed subjective reactions, in some cases using participants’ visceral and raw language, for how those experiences made them feel. Much—but not all—of the data from the RAP interviews were used to develop the survey items and q-sort statements used at later stages of my dissertation study. I made the decision to analyze only the interview data for specific perspectives on the interaction between teachers and evaluators for providing insight to what teachers’ felt about evaluators’ approaches and how those approaches impact the efficacy of the evaluation feedback conference.

Limitations

The overarching intent of the research apprenticeship study was to capture representative teachers’ voices about their experiences with the way their districts (namely administrators who are lead evaluators) are implementing teacher evaluation processes, primarily the evaluation feedback conferences and interactions during that conference. The number of teachers

interviewed is only a very small sampling of teachers who are experiencing the implementation of New York State Education Law §3012-d (APPR) evaluation system. This small sampling also is limited to Caucasian female teachers, which may have influenced the perceptions of the process. This demographic limitation has an impact on the findings from the interviews related to unarticulated issues of inter-personal relationships with evaluators based on race, ethnicity, and gender/orientation or on male-female ratio in the administrative ranks responsible for evaluations. A larger and broader representation of teachers of various levels and teaching contexts, accounting for variations in the ratios of teachers/lead evaluators by race/ethnicity and gender orientation/sex, is necessary to understand how personal demographics may impact the generalizability of how the teachers perceives the inter-personal conflicts and tension they experienced in the reported conferences.

A limitation related to this qualitative research approach is the un-fixed nature of narrative itself (Beuving & de Vries, 2015; Fraser & Jarldorn, 2015; Josselson, 2007; Phillips, 1997). Narrative discourse analysis, as an ethnographic/qualitative research method, puts into question the empirical truth and control of the narrative for understanding complex social constructions from participants' perspectives, which can border on self-indulgent and narcissistic (Chase, 2005; Polkinghorne, 2007). This perception of the data has the potential to shift the impact of the narratives-as-data, from being a means to empower the participants to push-back against prevailing and marginalizing narratives to a position that is too subjective and personal to be taken seriously as informing any social action (Beuving & de Vries, 2015; Fraser & Jarldorn, 2015; Polkinghorne, 2007). It was my responsibility as the researcher to ensure that the voices came through the reporting of data with care and consideration for representing participants'

authentic sense-making so they would be empowered to take action and not be further marginalized through my lack of mindfulness when reporting that data.

A contextual limitation includes other inter-personal relationships such as the evaluators' experience conducting feedback conferences, experience level of teachers being evaluated, and teacher-evaluator relationships not related to evaluations, all of which may have implications for how APPR conferences would be conducted. That being said, the limited sampling of teachers at the first stage of this study included a range of teaching experience (3 years to 30), range in representative districts (rural-suburban-urban) and range of grade/content levels (elementary, middle school science and math, high school science and ELA). Even this small sample of teachers revealed a range of evaluation experiences with multiple evaluators across and between their districts during their teaching careers that allowed for the range of interpretations and understanding teachers' perceptions of APPR feedback experiences.

There are also limitations related to how the participants were not asked to consider or give perceptions about how student assessment or growth scores were included as part of the APPR process. The narrative data does not consider how the inclusion of those scores would impact teachers' perceptions of the over-all APPR process. I intentionally ignored the topic of student test scores because it was not the focus of the research questions for this portion of the study. While some of the participants mentioned the scores, it was my direct decision not to include prompts that would explore or redirect participants to expanding on the topic as related to their feedback/post-observation conference.

Quantitative Paradigm: Survey Research

The decision to use a survey for the quantitative approach came about because I needed a way to reach a larger number of participants to explore the issues of evaluation feedback

effectiveness-utility from the position of the teachers. Survey research allows teachers to be informed observers, describing their experiences of how evaluators are engaging them in feedback conferences as part of their APPR evaluation. As observers of their own experiences, teachers are positioned as the ‘expert’ on their own experiences so they would be able to provide a rendering of their individual evaluative context. The data collected from this approach would contribute to understanding how teachers perceive the various ways evaluators conduct feedback conferences and how effective or ineffective they perceive those conferences to be over-all.

The survey items merged the themes from the RAP interviews with prominent themes from research literature on educational supervision and evaluation (see Table 1 in Chapter 2), drafts of a SOE Teacher-Leadership Simulation rubric, and a previous Syracuse University Teacher Evaluation Survey (see Appendix B: Syracuse University Study Council Teacher Evaluation Survey, 1995) that aligned with the dissertation research questions. Following a cyclical process for developing survey questions and items (Alreck & Settle, 1995; Babbie, 1990; Fowler, 1995, 2014; Onwuegbuzie, et al., 2010), the qualitative findings from the semi-structured interviews were used to generate quantitative survey questions. The themes went through three cycles of conceptual operationalization, discussing and defining indicators, and revising indicators (Luyt, 2012) which eventually became the survey statements describing the feedback process, as well as contextual factors that potentially affect the process (Luyt, 2012; Onwuegbuzie, et al., 2010; Teddlie & Tashakkori, 2006).

Using this collected data and information, the sections of the survey were deliberately constructed and sequenced as a way for participants to objectively report on evaluators’ behaviors when conducting evaluation conferences. There was considerable attention paid to the construction of the survey items to ensure the prompts and items genuinely represent individual

and varied experiences with how evaluators conduct conferences across and between districts. Another important consideration was on how the items would be perceived, such that participants would view the items as asking about their experiences and not as asking them to ‘tattle’ on administrators-evaluators. The survey design specifically framed the question, prompts and items in terms of objective behaviors that participants may have observed and could report on (Babbie, 1990; Fowler, 2014). By doing so, the survey invited teachers to think of themselves as objective observers as a way to minimize participants’ subjectivity. The sequence of the survey sections was also specifically ordered such that participants would respond to objective items/sections before responding to a last set of more subjective items/prompts that asked them to assess the effectiveness of the feedback conference. The purpose of this sequence was meant to afford participants the opportunity to respond objectively to items about the specific approaches under study before having them respond subjectively about the over-all experience (see Appendix C: Teacher Evaluation Conference Survey, 2019).

To enhance specificity and minimize subjectivity, the introduction of the survey asked participants to focus on just the most recent (within the year) evaluation conference. The decision to have participants focus on the most current conference experience was based on an underlying theme that emerged from RAP interview data and the most current research literature on teacher evaluation conferences. Interview participants and the literature note how evaluators’ approaches and protocols for conducting conferences seem to be changing or shifting from previous approaches; interview participants, specifically, describe multiple-varied experiences across their careers between and within districts, especially when they compared evaluation processes before and after the APPR law. I assumed many of the survey participants would have similar varied experiences with evaluation conferences. The implied underlying cause of this

change/shift seems to be related to the implementation of the state-mandated APPR law. This underlying theme is important to explore, but I decided to limit the scope of the survey to the current conference experience so that participants would only need to ‘remember’ a current experience rather than give a ‘composite’ report of multiple experiences.

To further encourage objectivity, the items were written as depictions of observable behaviors of the evaluators rather than as assessments about those behaviors. The survey’s design was intended to collect quantitative data from teachers’ rating of statements, along Likert scales ranging from “very accurate” to “very inaccurate,” on dimensions associated with feedback conferences. The dimensions include context of conferences, topics discussed as part of the conference, general approaches of evaluators when conducting feedback conferences, how evaluators provide for teachers' involvement in the feedback conference, and the general framework of feedback conferences that make them effective for the teachers. Writing the items this way would minimize interpretation of the items so the focus becomes reporting, not commenting, on the evaluators’ behaviors. The use of an “accuracy” scale (i.e. “very accurate”—“tends to be accurate”—“as accurate as not”—“tends to be inaccurate”—“very inaccurate”—“Do not know”) for describing the general and interpersonal approaches used by evaluators, and a scale for the extent (i.e. “Quite a lot”—“a fair amount”—“to some extent”—“not much”—“not at all”—“do not know/cannot recall”) to which the evaluators discussed general topics during the conference allowed participants to differentiate their experiences for what they know to be ‘accurate’ and ‘to what extent’ when evaluators conducted conferences in relation to their specific experiences, more so than if a scale of ‘agree/disagree’ were used.

The option for “Do Not Know/Cannot Recall” for the sections rating accuracy, with the exception for items on general situations and relationships which did not have this option, was

meant to be a different way to give a ‘non-response’ other than skipping the question. The option to not respond to prompts or items was extended in the invitation to participate, and then again in the informed consent statement at the beginning of the survey. Including this response option aligns with survey research methods as a way to make the survey more appealing in case participants felt pressure to complete the survey as an administrative directive, even with assurances of anonymity and confidentiality of participants’ responses when reported.

A second level of limited qualitative data were collected at the end of the survey in two separate, open-ended constructed-text responses where teachers could expand on their experiences not addressed in the previous statements. The constructed response prompts were intentionally added to the end of the survey to offer participants the opportunity to articulate thoughts on feedback conferences over-all which may not have been represented in the previous survey items. From the participants’ position, the constructed responses allowed their voices to be heard in their own words; from my position as the researcher, the constructed responses allowed for more data to be collected that could also be interpreted to support the findings from the previous research project as well as shed light on the quantitative correlations between the dimensions of effective feedback conferences and final set of items on effectiveness (Babbie, 1990; Fowler, 2014; Luyt, 2012; Onwuegbuzie, et al., 2010; Teddlie & Tashakkori, 2006).

Once the survey was in Qualtrics (2019), a survey program licensed by Syracuse University for use of approved research projects, and the project was approved by Syracuse University’s IRB, the survey was administered to a random stratified systematic sample with a random start of New York State (NYS) public school buildings. The list of public-school buildings came from the New York State Education Department’s directory of schools, available as an excel file on its website (NYSED, 2018), which listed a total of 7,304 buildings across 68

counties at the time of access. The sample frame was reduced from this initial list to 2,850 buildings, excluding New York City schools and my district. The sample frame excluded New York City schools because of the complexity for obtaining permission to participate from the governing school board, chancellor and regional superintendents; my district was excluded to avoid potential skewing of data because of bias on the part of colleagues knowing my research and conflicts of interest with the district administration about my research. The sample frame was then stratified alphabetically by NYS county and then by individual district/building BEDS codes. This stratification was done to obtain a more representative sample across the state. Using a sample ratio of 1:88, a sample of 100 schools were identified from the sample frame. Even without New York City Schools, this sample allowed for gathering impressions from a large population on multiple dimensions about the research topic and addressed all three research questions (Creswell, 2014; Fowler, 2014; Leech & Onwuegbuzie, 2009).

The sample population were members of the faculty employed in NYS public school buildings of the sample frame (100 school buildings) who have had a performance evaluation feedback conference experience within the current or previous school year. Data were collected from the sample population with an online, self-administered questionnaire through the school districts' internet network, with permission of school district leadership (i.e. superintendents and principals), via teachers' school email accounts. To reach this population, the superintendents of the selected buildings were contacted initially by email and post invitations to participate; the invitation to participate also included a copy of the survey so the superintendents would be aware of the prompts and items teachers would be rating.

When a superintendent indicated interest to participate, a paper-post letter was sent to the superintendent with specific information for coordinating the survey information (i.e. Qualtrics

URL link to the survey) with the district's IT department. All leaders in participating schools were sent a paper invitation explaining the survey, then a follow-up email invitation with an URL link for the individual building's survey. I made the decision to use Qualtrics to distribute the survey and collect data over other survey software programs because I wanted to ensure as much anonymity for the identities of all participants, and the Qualtrics program allows a level of anonymity for participants that other, less sophisticated survey instruments may not afford. In their invitation, teachers were informed that their participation was completely voluntary and assured that their responses would be kept anonymous and confidential as well (see Appendix D: Dissertation Study IRB Materials). Responses from the faculty of individual school buildings were collected in Qualtrics as distinct data files, but no other distinguishing school feature connected the responses to the participants.

A total of 98 invitations to participate in the survey were sent out to district superintendents (superintendents of two buildings within the same large district were sent a single invitation); five superintendents (representing five separate public school districts) agreed to participate. With the limited number of participating districts, the district demographics represent a narrow population. The five districts identify as predominately suburban, middle class to affluent, and rural; the survey did not ask participants to identify this demographic information.

Once agreeing to participate, the districts were given a window for responding; at the time when the response windows closed for all participating districts, one district did not have any participation and the other four districts had a combined total of 58 responses, of these 39 responses yielded complete surveys with usable data. The collected surveys from the participating districts reflected a limited range of professional demographics since the survey

purposefully did not ask about personal demographics such as age or gender. The professional demographics also are limited to the number of years teaching and in what capacity or context they were observed by an evaluator for purposes of an APPR evaluation. Since the survey was presented as confidential and anonymous, there was no way to follow-up with participants to ask more identifying professional demographics. As the researcher, I know which survey responses can be attributed to which participating district, but this information only will be used for a generalized follow-up report to all five superintendents at the conclusion of this study.

The raw data were transferred from Qualtrics as Excel data files and uploaded into a statistical software program, SPSS (2019), creating one combined data file. Once uploaded to SPSS, the individual survey item stems were revised for conciseness to perform factor analysis and rotation, and all participant response values were reviewed to ensure accuracy and completeness. The collected responses also were put through a data-cleaning process (Babbie, 1990; Fowler, 2014) that included deleting incomplete surveys and coding any missing values for individual items in the remaining surveys as “mean” for purposes of factor analysis, rotation, and correlation (Babbie, 1990; Fowler, 2014; Osborne, 2014).

Even though the over-all response to the survey did not yield as many participants as predicted, enough responses were completed to allow for statistical analysis of what data were collected. The analysis process I used followed the precepts of Exploratory Factor Analysis. Specifically, Exploratory Factor Analysis (EFA) is a refined, manageable way of determining a statistical model that will explore relations among the items (DiStefano, Zhu, & Mîndrilă, 2009; Osborne, 2014; Peterson, 2017; Reio & Shuck, 2015) and uncover underlying variables “without imposing an a priori structure on the factors” (Reio & Shuck, 2015, pg. 13). As a process, EFA

required multiple iterations of factor analysis using multiple variations across the six sections of survey items (Babbie, 1990).

I used SPSS to apply EFA to the raw survey data (excluding the text from the qualitative, constructed responses) because the program offers multiple factor extraction and rotation options. I employed Principal Axis Factoring (PAF) with an Oblimin Rotation (Kaiser Normalization) to determine which items, across and within sections of the survey, would factor significantly. PAF is most appropriate for smaller samples with problematic outliers, and Oblimin Rotation/Kaiser Normalization minimizes the number of factors while also maximizing the intercorrelations of variables within each factor (Osborne, 2014). The SPSS output (report of calculations for inputted variables) after performing a PAF/Oblimin-Kaiser rotation of data included both pattern and structure matrices; for purposes of determining factors I used the structure matrix for each set of survey items/sections because of the simplicity of the correlations. Based on the factor loadings in the structure matrix and performing Cronbach's Alpha Reliability checks (Osborne, 2014), I determined which items to finally include for each factor cluster for purposes of identifying variables and interpreting factor correlations. I used the same sets of matrices and scales for each of the factor clusters. The following tables provide the summary of factor loadings, means, standard deviations, and Cronbach's Alphas performed on the sections of items from the survey; discussion of items included or excluded from factors follow each of the tables.

Table 4A

Summary of Exploratory Factor Analysis Results for General Approaches Evaluators Use When Conducting Conferences (N = 39)

Item	Factor Loadings		
	Identifies and Addresses Evaluator's Concerns	Identifies and Addresses Teacher's Accomplishments/Strengths	Support for Collaboration/Follow up
Discussed topics before conference	.373	.056	.376
Explained goals for conference	.294	.300	.445
Discussed evaluation rubric elements	.140	.450	.305
Discussed limited rubric elements	-.003	-.010	-.020
Discussed evaluator's ratings	.071	.345	.690
Explained rationale for ratings	.143	.398	.644
Expressed appreciation for teacher	.092	.908	.256
Discussed teacher's general accomplishments/strengths	.044	.950	.320
Discussed concrete examples of accomplishments/strengths	.237	.887	.283
Discussed ways to build on accomplishments/strengths	.440	.582	.570
Discussed concerns about practice	.805	.183	.222
Discussed concrete examples of concerns	.899	.066	.283
Discussed what actions to take to address concerns	.959	.181	.458
Discussed concerns/actions are highest priority	.792	.044	.618
Encouraged collaboration with other teachers	.494	.132	.779
Encouraged agreement on steps teacher will take post conference	.575	.102	.735
Discussed evaluator's actions to support teacher's improvement	.671	.025	.761

	Identifies and Addresses Evaluator's Concerns	Identifies and Addresses Teacher's Accomplishments/ Strengths	Support for Collaboration/ Follow up
<i>M</i>	3.591	1.737	2.903
<i>SD</i>	1.285	0.842	1.067
α	.955	.873	.903

Note: Structure Matrix for Principal Axis Factoring/ Oblimin with Kaiser Normalization
Highest loading for each item highlighted, provided loading is .500 or greater. Item with nearly identical loadings >.500 assigned to Factor 2 on conceptual grounds.

Based on the factor loadings for this section of the survey, two independent variables emerged; two items not assigned to any of the factors, “discussed evaluator’s ratings” and “explained rationale for ratings,” did not conceptually factor with any other items so therefore I excluded them in any of the factors including composite factors. I assigned the last three items to a composite factor, and did not assign the first to any factor at all despite showing some reliability as a two-item factor; I felt the factors I was able to extract from this section of the survey sufficed and this two-item factor did not contribute conceptually to the over-all interpretation of the results.

Table 4B

Item	Factor Loadings	
	Seeking Teacher's Input	Attention to Teacher's Input
Asked for teacher's input on what to observe	.788	-.408
Asked for teacher's input on conference focus	.865	-.338
Urged the teacher to identify goals and/or concerns to discuss	.826	-.441
Encouraged the teacher to provide information on topics being discussed	.759	-.637

Urged the teacher to identify anything that might affect the teacher's ability to teach effectively	.752	-.346
Was open to teacher's opinions, even if different from evaluator's own	.442	-.906
Paid close attention to what the teacher had to say	.491	-.958
Used strategies like paraphrasing, maintaining eye contact and other non-verbal cues to convey attention to teacher's opinions	.465	-.767
Used open-ended questions that invited discussion rather than assertions that would close off discussion	.417	-.889
Encouraged the teacher to suggest options for addressing identified concerns	.599	-.592
Invited the teacher to suggest options for addressing identified concerns	.704	-.466
Emphasized the need for the teacher and evaluator to reach conclusions both could support	.660	-.340
<i>M</i>	2.730	1.734
<i>SD</i>	1.053	1.025
α	.915	.929

Note: Structure Matrix for Principal Axis Factoring/ Oblimin with Kaiser Normalization
Highest loading for each item highlighted, provided loading is .500 or greater. Item with nearly identical loadings >.500 assigned to Factor 1 on conceptual grounds.

The section of the survey on the approaches evaluators use to involve teachers in the conference contributes two complete factors. The notable item, "Encouraged the teacher to suggest options for addressing identified concerns," loaded almost equally between the two factors (.599 and -.592 respectfully); I assigned this item to the first factor since the approach this item focuses on conceptually fits better with the other items that loaded on this factor. The seven items on this factor all exhibit behaviors that directly involved the evaluator to engage the teacher, while the behaviors of the other factor reflect the way the evaluators would pay attention to what the teacher is saying.

Table 4C

Summary of Exploratory Factor Analysis Results for the Extent of Focus by Evaluators on Specific Topics During Conferences (N = 39)

Item	Factor Loadings		
	Teacher's Practice	Summative Assessment	Teacher's Relationship with others
How the teacher plans lessons	.549	.294	-.556
Teacher's expectations for students	.709	-.125	-.371
Teacher's content knowledge	.604	.049	-.351
Whether and how teacher's students actively engaged in learning	.752	-.136	-.167
How the teacher assesses students' learning	.726	-.226	-.405
Teacher's classroom management and/or relationships with students	.612	.312	-.118
How the teacher differentiates instruction	.706	.205	-.358
Teacher's relationship with other teachers	.284	.196	-.765***
Teacher's interactions with parents/guardians/care-givers	.257	.564	-.616***
Teacher's plans for improving teaching	.377	.308	-.794*
Student test scores	.116	.657	-.339
Other data or information, besides test scores, on what students have learned	.607	.148	-.343
Whether students are meeting expected learning standards or objectives	.723	.370	-.125
Compliance with district/school policies	.280	.718	-.360
The ratings the evaluator assigns	.181	.699	-.229
<i>M</i>	2.201	2.673	***
<i>SD</i>	0.748	0.924	***
α	.872	.663	.645

Note: Structure Matrix for Principal Axis Factoring/ Oblimin with Kaiser Normalization

Highest loading for each item highlighted, provided loading is .500 or greater.

*Item with loading >.500 assigned to composite factor "Support for Collaboration/Follow up" (see Table 4A and Table 4D) on conceptual grounds.

********M* and *SD* were not determined for this 2-item factor

The items in the section of the survey on what topics evaluators focused on provided one significantly loaded factor and one two-item factor, “teacher’s relationship with others;” one item, “Student test scores,” did not factor with any other factor (either within this section or with other composite factors) and two items loaded with a composite factor, “Summative Assessment.”

Table 4D

Item	Factor Loadings		
	Fair Appraisal	Awareness of Teacher’s Practice	Support for Collaboration/ Follow up
Understands the curriculum teachers are responsible for following	.594	-.796	.433
Understands the instructional challenges teachers face	.549	-.676	.266
Observes teaching on a regular basis	.473	-.696	.455
Asks for teachers’ advice on issues that affect them	.496	-.865	.426
Conveys a clear vision of what she/he wants students and school to accomplish	.512	-.913	.489
Encourages teacher to turn to each other for advice	.218	-.387	.877
Encourages teachers to try new ideas, even if doing so might mean making mistakes	.673	-.782	.790
Takes time to give each individual evaluation careful attention	.764	-.752	.656
Appraises performance fairly	.947	-.659	.191
Shares opinions respectfully with teachers	.857	-.549	.291
Uses accurate information when discussing performance	.839	-.487	.117
Shows that he/she has the interests of students in mind	.828	-.508	.279
Shows an awareness of what teachers have done to improve teaching	.577	-.780	.175

Shows that she/he is a good judge of teachers' effectiveness	.839	-.763	.351
Uses strategies to help see situations teachers face from different vantage points	.769	-.660	.284
Gives useful perspectives on things he/she observes in teaching	.809	-.602	.489
Makes useful suggestions for specific things teachers might do to change teaching	.636	-.583	.645
<i>M</i>	1.691	2.263	2.903
<i>SD</i>	0.763	0.991	1.067
α	.941	.909	.903

Note: Structure Matrix for Principal Axis Factoring/ Oblimin with Kaiser Normalization
Highest loading for each item highlighted, provided loading is .500 or greater.

Items from two sections of the survey, on general approaches the evaluator took when conducting the conference and the general situations and relationships that affect conducting conferences, did not significantly load when factored separately, but did when factored together.

Along with two significant factors, three items loaded on a composite factor, "Support for Collaboration/ Follow up."

Table 4E

Summary of Exploratory Factor Analysis Results for Measures of Conference Effectiveness (N = 39)

Item	Factor Loadings		
	Cultivating Relationships	Strengthening Teaching	Accountability
Helping the teacher to improve knowledge and skills	.727	.865	
Helping the teacher make sense of problems/concerns teachers face in teaching/other work with students	.702	.908	
Helping teachers develop own solutions for addressing identified goals or concerns	.716	.862	
Providing the teacher with feedback that could be used to strengthen teaching/other work with students	.846	.638	

Providing recognition for the teacher's efforts	.874	.406	
Providing the teacher with an opportunity to reflect on own performance	.856	.448	
Providing an opportunity for serious discussions of different approaches to teaching/other work with students	.787	.800	
Ensuring teacher conform to district/school policies	.127	.415	
Identifying ways to further the teacher's professional development	.778	.699	
Arranging for the teacher to get help or resources to improve teaching/other work with students	.702	.570	
Fostering trust between teachers and administrators	.879	.412	
Holding teachers accountable for the teaching strategies used	.491		.889*
Holding teachers accountable for students' learning	.528		.771*
<i>M</i>	2.356	2.731	2.218
<i>SD</i>	0.996	1.183	0.985
α	.934	.879	.903

Note: Structure Matrix for Principal Axis Factoring/ Oblimin with Kaiser Normalization
Highest loading for each item highlighted, provided loading is .500 or greater.

*Item with loadings >.500 assigned as third/2-item factor on conceptual grounds.

The last section of the survey asked teachers to report on the effectiveness their conferences were for various aspects of their practice. Two strong and one two-item factor emerged; one item, “ensuring teachers conform to district/school policies,” did not load on any factor.

One full section from the survey asking about general Teacher-Teacher situations and relationships that affect conducting effective feedback conferences loaded significantly within

only this section as two factors; one item did not load with any other items in the section. Taken collectively as a section of outlier items, seven of the eight items loaded across two factors:

Collaboration on Instructional Approaches [reliability: .871]

1. agree that all students can meet high expectations
2. agree on what constitutes effective teaching (.855)—highest loading item for factor
3. regularly discuss curriculum and instructional issues with each other
4. participate in ongoing professional development with each other
5. coordinate their instruction with each other

Collaboration on Student Learning and Assessment [reliability: .686]

6. develop common approaches to assessment of students
7. work together to analyze data on student learning (.723)—highest loading item for factor

One, non-factoring item ([teachers] pay more attention to how test scores and other evidence of student learning affect their evaluation ratings than to administrators' suggestions based on observations) did not work with any of the items within this section, nor did it relate conceptually with other items; therefore, over-all I excluded this section of items from my analysis.

Using this information from SPSS, I identified nine independent variables related to evaluators' general approaches to conducting conferences, and topics addressed by evaluators during those feedback conferences:

- Identifies and Addresses Evaluator's Concerns
- Addresses Teacher's Accomplishments/Strengths
- Seeks Teacher's Input
- Attention to Teacher's Input
- Broad Discussion of Teachers' Practice
- Evaluator's Fairness
- Awareness of Teacher's Practice

- Summative Assessment of Practice
- Providing for Follow-up

Using the same procedure to determine factors and complete factor analysis, I identified three dependent variables that reflect measure of conference effectiveness:

- Effectiveness=Cultivating Teacher-Evaluator Relationships
- Effectiveness=Strengthening Teaching
- Effectiveness=Accountability

The surprising outcome of the survey data is how the factor clusters ended up showing high levels of reliability for a relatively small amount of raw data. Another notable outcome is how the data was enough to analyze the correlations between the independent and dependent variables, as well as perform a regression analysis to develop a composite model for predicting evaluation effectiveness based on three sets of effectiveness measures and four independent factors of evaluators approaches and topics. The next chapter provides more thorough discussion and analysis of the factors, reliability findings, correlations and regression model for predictions of conference effectiveness.

Limitations

There are notable limitations with the size of the sample frame and collecting limited demographic details from such a small sample frame. The small sample frame of 100 public school buildings represented approximately 1% of the number of public-school buildings in NYS. Since only five districts agreed to participate, and signed letters of consent to gather responses from the teachers in only one building in each district, the narrow range of district demographics to just suburban and rural has an impact on the generalizability of the findings across a larger population. The use of this demographic information is rendered moot in the

analysis process since the data collected from a limited district-demographic sample cannot be used to generalize across the whole target population.

With such a small sample frame, collecting limited demographic details presented the possibility of over/under representation across the years of practice and ‘current assignment’ in the surveyed district. The over-all small sample frame and limited demographic details raised concerns, prior to actually collecting survey data from the noted districts, about whether enough data would be generated to be sufficient to generalize across the entire population (Babbie, 1990; Fowler, 2014; Hedrick, et al., 1993). This concern shifted to whether enough data would be collected after a low number of districts agreed to participate, and then when a low number of teachers in those districts actually participated.

Another issue related to the low response rate was how the confidentiality/anonymity was structured in the study design. One person, the district superintendent, was contacted to agree to participate, and then I relied on those superintendents to distribute the survey to all the teaching faculty in the designated building. There was a possibility that issues of inter-personnel-district relationships, of which I was unaware, could cause miscommunication or tension between administration and teachers that would result in non/low responses. The underlying inter-personal-district relationships could be a contributing factor since teachers may have reacted to what they might have felt was a ‘directive’ by the superintendent/principal so there is a possibility they ignored the request or participated to skew the data.

Another related contributing cultural factor could be from participants’ lack of investment in the evaluation process as a whole, separate from how the district implements the process. Participants also may not have been comfortable with the format of online surveys, or surveys in general. The difficulty with completing the survey from school district email accounts

on personal digital devices (i.e. smart phones, iPads/tablets, home computers) or the reluctance of using personal digital devices may have also contributed to the low-non response rates.

To reduce non-responses, the survey design addressed the anonymity issue with a preliminary letter sent to each teacher by email and paper-distribution (to be placed in building mailboxes) that included the URL for accessing the survey, and the beginning of the survey provided the choice to participate as part of the consent statement. Another way to reduce non-responses was an appeal for teachers' input on a topic that has conventionally kept teachers' voices marginalized (Dillman, Smyth, & Christian, 2014). Because the survey assures anonymity, no post-mailings were used as follow-up for non-respondents, but superintendents were asked to send an email reminder to all faculty members before the close of the survey response window. Another important step to lessen the response bias was to ensure the survey instrument was well designed and addressed the issues that were important to the respondents (Dillman, et al., 2014; Fowler, 2014).

Mixed Paradigm: Q-Methodology—Focus Group Q-Sorts

Q-Methodology is a hybrid quant-qual research approach that can be used as a stand-alone mono-method, but also used in conjunction with other methods as part of an over-all larger mixed methods study such as this one. Q-methods/approaches are “a more interactive and entertaining way of engaging research subjects and drawing out their views...[that can] produce unexpected results” (Eden, et al., 2005, pg. 420). Using this approach, I engaged focus groups in an activity, as the last piece of the QUAL-QUANT-MIXED research design, that collected participants' perspectives on evaluation feedback conferences in a different way than through interviews and a survey.

Whether Q-Methodology is used as a mono-method or as one component of a larger study, either usage requires the conscious, reflexive use of the method for making valid and important contributions to the study that other methods do not/cannot provide (Dziopa & Ahern, 2011; Eden et al., 2005; Ellingsen, et al., 2010; McKeown & Thomas, 2014; Ramlo, 2016; Watts & Stenner, 2005; Wright, 2013). As a ‘mixed-hybrid’ method, Q-approach has distinct processes and phases that combine quantitative and qualitative components (Eden et al., 2005; Ellingsen, et al., 2010; McKeown & Thomas, 2014; Ramlo, 2016; Watts & Stenner, 2005; Wright, 2013); qualitative data on participants’ self-referential, subjective perceptions are simultaneously collected with quantitative data that can be used for statistical analysis of that subjectivity (McKeown & Thomas, 2014). Q Methodology blends the empirical nature of objective, quantitative factor analysis (R Methodology) with the humanistic, qualitative identification of representative themes to “measure individuals’ affinity with [shared] views, as well as similarities and divergences amongst individuals” (Eden, et al., 2005, p. 414) on those same shared views (Ellingsen, et al., 2010; McKeown & Thomas, 2014; Ramlo, 2016; Shemmings, 2006; Watts & Stenner, 2005).

In much the same way as interview and survey construction is an iterative process, a Q-approach also requires a process for constructing the specific components that make the approach ‘Q.’ As with other research instrument development, a Q-approach requires first creating a *Concourse*, which is the process of identifying and then collecting information on a research topic, current issues related to the topic, and existing (and potential) points of view related to the topic. The concourse for my study was based on my RAP interview response data, survey factors and items with high reliability and correlation to conference effectiveness

measures, constructed responses from the survey, and themes from my literature review. The concourse is used to inform the next stage of instrument design.

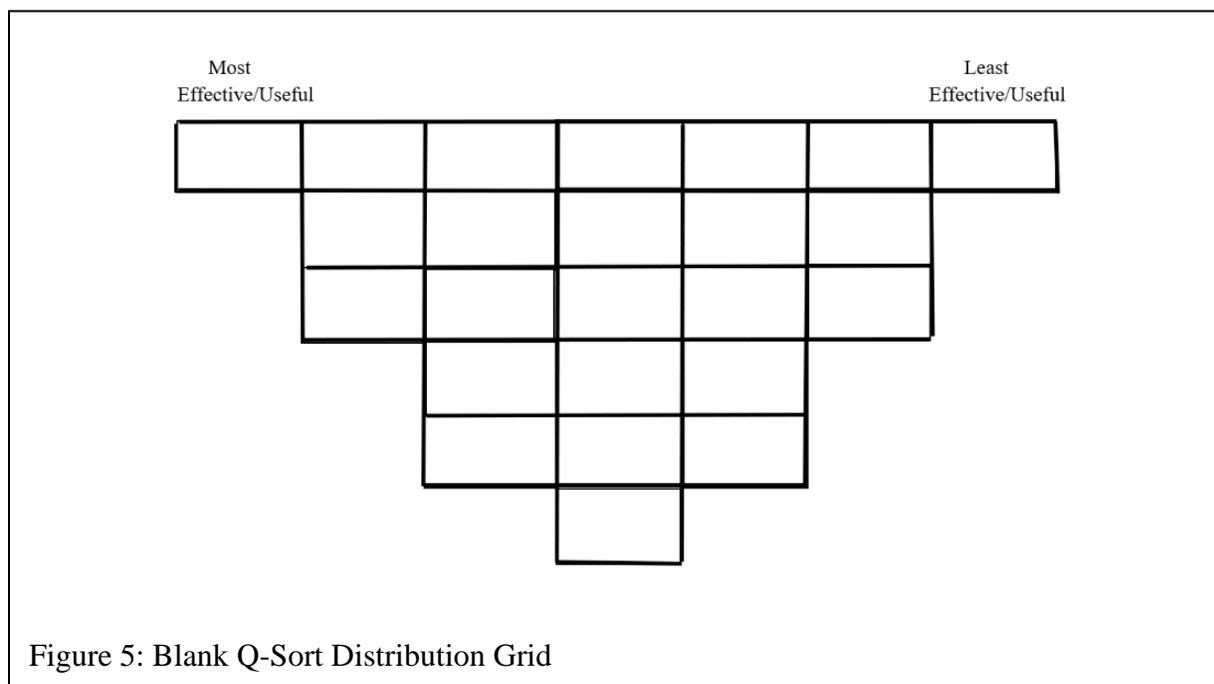
From the concourse, a *Q-Sample* is developed to represent multiple facets and perspectives of the concourse in the form of a comprehensive collection of statements. The Q-Sample should have a limited number of statements to be manageable for the next step of the process (Q-sorting and statistical analysis), but still have enough statements that will capture participants' perspectives as fully as possible. When determining my Q-Sample, I started with over 30 representative statements for two guiding prompts; one prompt represented the independent factor variables and the other represented the dependent factor variables from the survey factor analysis. I revised this Q-Sample because the two sets of statements and prompts became too complicated and could possibly confuse participants with over-lapping themes/variables. The final Q-Sample became a total of 24 statements:

1. The evaluator has an idea of what I teach and how I teach in order to discuss his/her concerns, how I will address those concerns, and what I will give the highest priority.
2. The evaluator lets me identify aspects of my teaching that I consider to be areas of concern, how to address those concerns, and what priority I should address those concerns.
3. The evaluator lets me know before the conference anything he/she wants to discuss and explains what he/she wants to accomplish in our conference.
4. The evaluator asks me about what to look for in the observation.
5. The evaluator discusses all elements of the evaluation rubric the district uses.
6. The evaluator discusses only the elements of the rubric she/he thinks are important/relevant to my evaluation rating.
7. The evaluator discusses and explains some/all of the ratings the she/he plans on giving me.
8. The evaluator encourages me to suggest options for addressing both of our concerns as well as collaborate on coming to consensus on how to address concerns both of us can support.

9. The evaluator is open to my opinions, even when they differ from his/her own, which shows he/she is paying attention to what I have to say about my practice.
10. The evaluator takes the time to give my evaluation individual attention so that the discussion focuses on useful feedback.
11. The evaluator discusses whether/how my students are actively engaged in learning/ meeting learning objectives and how I assess their learning.
12. The evaluator focuses the conference on how I plan lessons/adapt instruction for different students.
13. The evaluator asks about how I manage my classroom and student issues and interactions with parents/care-givers/guardians (i.e. discipline/positive interactions, contact logs, phone calls, etc) to help make sense of problems/concerns these relationships present in my teaching or other work with students.
14. The evaluator encourages me to develop relationships with other teachers who share the same concerns as I do or seek out other teachers for advice on how to address some of the concerns/issues brought up in the conference.
15. The evaluator focuses on/discusses in-depth useful suggestions for specific actions I might take to change my teaching, including new ideas that may mean making mistakes, with what support he/she will give me based on what we decide I will try to do.
16. The evaluator asks me to come to the conference prepared to discuss anything I think needs attention, including my goals, concerns and/or something from the observation.
17. The evaluator discusses my expectations for student learning.
18. The evaluator discusses how students score on assessments of growth, and how I use student information in practice.
19. The evaluator genuinely expresses appreciation for the work I do and acknowledges my accomplishments and strengths that I can build on to improve my practice.
20. The evaluator is respectful sharing insights/opinions of teaching and asks for advice on how to address issues that affect me.
21. The evaluator uses accurate information when discussing what was observed and shows an awareness of what I have done to improve teaching.
22. The evaluator shows that he/she is a good judge of my effectiveness because he/she regularly observes my teaching, understands the curriculum I follow and understands instructional challenges I face.
23. The evaluator conveys a clear vision of what she/he wants students and the school to accomplish, showing that she/he has the interests of students in mind.

24. The evaluator focuses on using the conference to help me see situations related to my teaching from different vantage points, offers useful perspectives on things she/he observed in my teaching, and identifies ways to further my professional development on those situations and my over-all practice.

The next stage in the Q-approach is for participants to ‘sort’ or rank-order, called *Q-Sorting*, the q-sample statements onto a distribution grid. To complete the sorting process, each participant is given a set of the statements on 3” x 5” index cards and instructed to put the statement numbers into a distribution grid, printed on 8.5” x 14” paper, according to how each participant felt the statement addressed the over-arching prompt. The grid follows a quasi-normal distribution pattern with two poles, from “most effective/useful” to “least effective/useful,” which is purposefully designed to prevent participants from making casual choices in distributing the statements across the grid, under neutral, or on just one side (Wright, 2013).



The multiple columns of the grid, with some columns having multiple rows, forces participants to make choices that best reflect their perspectives and also allows participants to have duplicate choices under multiple column headings, especially neutral. With such a distribution grid,

participants are ‘forced’ into making distinctions, which is useful in the analysis and interpretation stages of the approach for purposes of identifying what particular approaches participants seem to be most sensitive to for how evaluators conduct conferences.

Following the procedures noted in Q-Methodology study design, each participant (i.e. members of my focus groups) was instructed to consider an over-arching sort-prompt, “Based on your own experiences, how effective/useful are the general ways evaluators conduct conferences?” before assigning the statements into the grid. This over-arching prompt used the underlying research issue of how teachers are experiencing the evaluation feedback conference in the over-all APPR process. As a focusing question, the prompt gave participants a way to put the statements into context and put ‘themselves’ into the statements. Both the focusing prompt and how participants sort the statements under that prompt were meant to collect the subjective side of how teachers feel about evaluators’ approaches that make conferences effective. The resulting q-sort data from this process are different from qualitative discourse analysis data and quantitative survey data for how the q-sorts will show patterns within and between individual experiences, but also show differences between groups of participants for how they are individually sensitive to approaches evaluators use in feedback conferences.

The participants were given as much time as needed to read/comment on the statement cards, assign statements on the grid with written comments, and ask questions if needed. The groups’ discussion was recorded (with consent), and all participants were encouraged to record (write) individual thoughts about the statements on the cards as well as the reasoning for placement of statements in the grid on the paper. Throughout the duration of Q-sorting, I encouraged participants to ‘think aloud’ about the statements as they sorted them and to verbally or textually (i.e. write down on paper) explain their reasons for positioning or assigning some

statements at either end specifically, as well as indications for why other items are assigned a more neutral value in the middle. At a time when all members of the group indicated they completed the sorting and responding, the whole group discussed the topic, shared (when-if comfortable) statement assignments and general impressions of the activity as beneficial for their own understanding of the topic. Participants were asked to complete an activity evaluation (much like workshop or class evaluations) as feedback on the topic and instrument (q-sort grid).

Selecting participants for the focus groups, called a *P-Set* in Q-research, was a deliberate process. The P-set should include participants whose experiences are the subject of the research study, as well as include “enough participants to define a factor which can be readily compared with other factors extracted from the data” (Wright, 2013, pg. 154). Since I relied on many of my teacher-colleagues thus far for interviews and other research projects related to coursework on this topic, I purposefully did not ask those same teachers to be included in the P-Set. Colleagues who participated in the previous research activities would have prior knowledge of my position on the topic, as well as my having already collected their ‘unvarnished’ subjective perceptions on this topic. I briefly considered the idea to include administrators in the q-sorting, with other teachers or as their own P-Set, but the specific focus of the research study is to gather *teachers’* perceptions about evaluators’ approaches. My intent with gathering other teachers for the P-Set was to collect data that I have not had the chance to hear or analyze until now that would include reflections/responses to previously collected responses on opinions and relationships established by earlier stages (i.e. interviews/survey) of the study.

To avoid over-lapping participants, I used a convenience/word-of-mouth method to gather teachers of various experience levels, grade levels, content areas and districts. I contacted participants via phone messages and with face-to-face requests. I was mindful of the limited

personal demographics of the interview sample, so I purposefully contacted participants who identified as male and range of ages, for a total of five of the 13 participants. The group over-all represented a range of professional demographics related to content area (i.e. Business Education, English, ESOL, General Education, Mathematics, Music, Social Studies, and Special Education), grade level (i.e. elementary, middle school, high school), years of teaching (i.e. first year to 33 years), and teaching context/district (i.e. urban, rural, suburban, BOCES).

I was able to convene two small focus groups, one a P-set of nine experienced teachers (including myself) and the other P-Set of four untenured teachers. Both groups convened at a neutral site for 90 minutes, and participants signed consent forms that informed them the sessions would be recorded for field-note purposes. I voice recorded the larger group on a personal recording device in sight of all participants, and recorded fieldnotes as a participant-observer with the second, smaller group.

After both focus groups completed the q-sorting process, I entered the individual q-sorts into a software program, PQMethod (Schmolck, 2014), to perform a statistical analysis akin to a conventional factor analysis. The factor analysis performed by PQMethod is different from traditional analyses usually performed by SPSS in the way Q-sorts are by-person factored rather than item factored. The PQMethod performs a Principal Component Analysis on the uncorrelated, individual q-sorts and then provides eight unrotated factors (the program's default number) with Eigenvalues for each factor. Based on the Eigenvalues, I chose to employ a Varimax rotation on only two factors so that each q-sort was associated with a factor. In this way, each participant's perspective is taken into consideration when factored, and I would be able to make interpretative decisions on how the factors represented the statistically significant similarities and differences among the subjective perspectives (Eden et al., 2005; Ellingsen, et

al., 2010; McKeown & Thomas, 2014; Watts & Stenner, 2005; Wright, 2013). Each factor, thus, represents the statements that a particular group of participants considered most significant. If only one factor were identified, the conclusion would be that there are no significant differences in participants' subjective perceptions. This factoring process represents the quantitative component of Q-methodology, while the comments from the recorded/observed discussions and individual, narrative written comments represent the qualitative component; taken as a whole, the combined q-sort data provided representative perspectives of the participants that are statistically rigorous, valid, and reliable.

Limitations

The philosophical assumptions associated with subjectivity (Stephenson, 2014) calls any Q-study's internal and external validity into question. The over-arching limitation of Q-Methodology, that being whether or not subjectivity can really be measured with validity and reliability, impacts how Q-data are used and for what purpose. Specifically, the subjectivity of the small sample P-set cannot be generalized across a larger population nor be inclusive of all possible perspectives that a larger sample would be able to represent (Wright, 2013). Even with the mindful way for who I asked to be included in this research method, this limitation of the narrow over-all P-set impacts how the data can be interpreted to provide insights on how the whole of the population may or may not align with the viewpoints of the included P-set (Kampen & Tamás, 2014). To address this limitation of lack of generalizability of Q-data, I used other approaches to corroborate, qualify or challenge inferences made with Q-data, and likewise Q-data shed light on inferences made with data from the other approaches.

The concourse also presents a limitation in that there is no concrete way to ensure the included statements of the concourse are complete and representative of the whole topic, issue,

and existing variations of perspectives. The dynamic nature of the independent variables and individual contexts that participants experience related to evaluation conferences (or the evaluation process as a whole) make complete representativeness elusive. The way to address this limitation is to be aware of the boundaries of the discourse for how it specifically addresses the research questions (Kampen & Tamás, 2014), thus avoiding including tangents not related to the topic and/or under-representation of the topic in the q-sort statements.

There are different ways to conduct the q-sorting with participants. One way is to allow participants to respond to given statements/sorting artifacts (i.e. visual prompts or music) in an unstructured format (i.e. no pre-determined grid), such as sorting into piles, that can be deconstructed and reconstructed depending on the focusing sort-prompt (McKeown & Thomas, 2013). Applying q-sort procedures in this way would allow the strength of participants' viewpoints to organically emerge and shift to reveal multiple layers of individual perspectives (Eden, et al., 2005; Stephenson, 2014). A more commonly used approach implements the sorting process with a narrow P-set (i.e. only teachers who have had evaluation conferences as recently as the previous school year) and quasi-normal distribution grid with a pre-determined number of cells to represent levels between most-to-least and only one cell for the poles of the grid (Eden, et al., 2005; Dziopa & Ahern, 2011; Shemmings, 2006; Watts & Stenner, 2005). Using the quasi-normal distribution grid and a narrow P-set helps "to differentiate nuances in different statements" which brings participants' subjective thinking into focus on the specific topic (Ellingsen, et al., 2010). I chose to implement the later approach in this study because the sorting procedure would use pre-determined statements derived from data collected from two previous groups of participants on the phenomenon under study, and the results from their sorting would contribute to confirming or disputing that data.

As a related limitation to altering the traditional Q approach is how I provided participants a structured, quasi-normal distribution grid for the sorting the statements. The literature on Q Methodology is inconsistent when discussing the issue of ‘forced’ (i.e. quasi-normal grid) versus an ‘unstructured’ approach to completing a q-sort (Eden, et al., 2005; Ellingsen, et al., 2010; Kampen & Tamás, 2014; Ramlo, 2016; Watts & Stenner, 2005); even with the inconsistencies around the issue of which distribution grid to use for q-sorting, Q Methodologists note that both approaches provide equal validity for the statistical analysis of the factor arrays which result from q-sorts (Dziopa & Ahern, 2011; Eden, et al., 2005; Kampen & Tamás, 2014; Shemmings, 2006). As the researcher, I focused on using the ‘forced’ quasi-normal approach since I felt the phenomenon under study would provide statistically stronger factor arrays from the limited participants in the focus groups (Dziopa & Ahern, 2011; Shemmings, 2006; Wright, 2013). However, there are Q Methodologists who consider using a more structured grid for q-sorting may not reflect how participants’ perspectives can shift or change depending on how they interpret the statements or for how they feel about the experience at various points in time. As a limitation, this forced ranking can impact the number of factors, thus limiting the diversity of the viewpoints and the generalizability of factors that do emerge (Kampen & Tamás, 2014). I was aware of this issue and how it may impact the validity of my findings, so I encouraged participants to ‘think-aloud’ so that I could capture as much as possible any ‘thinking-through-the-statements’ process-comments, including any shifts or changes in interpretation, to address this limitation.

Mixed Methods Validity

An internal challenge of Mixed Methods research relates to how experiences of participants are collected, interpreted, and reported. The design of the study must consider the

data content, interpretation, and findings for how the participants' reality (or 'truth') is rendered for representing their experiences. Whether the approach is qualitative, quantitative or both, truth is both individually and socially constructed and situated in the way understanding and reflection are organized and known (Gubrium & Holstein, 1997, 2000; Polkinghorne, 2007). No matter how many methods or tools are used, "[o]bjective reality can never be captured" (Denzin & Lincoln, 2000, p. 5), only understood through its representation. The Mixed Methods researcher, therefore, must be vigilant at all stages of the study to ensure that collection, content, inferences and reporting of data are continuously positioned in ways that not only represent the researcher's understanding of the phenomenon but also represents all participants' truths of their experiences to the best of the researcher's ability (Fine, et al., 2000; Long, 2017; Newman, Lim, & Pineda, 2013; Onwuegbuzie & Johnson, 2006; Onwuegbuzie, et al., 2011).

Some of the criticisms with Mixed Methods research focus on validity, reliability, and credibility issues related to paradigmatic differences, interwoven philosophical/theoretical frameworks, methodological weaknesses, and proficiency of using multiple methods (Cameron, 2011). Researchers who favor monomethod research approaches often express concern that mixing qualitative and quantitative paradigms creates contradictions and competing research agendas. To address these criticisms, the researchers Onwuegbuzie and Johnson (2006) and Onwuegbuzie, Johnson and Collins (2011) make the pragmatic response that "to search for workable solutions through the practice of research" may seem contradictory but will "enable one more fully to see his or her world" (Onwuegbuzie & Johnson, 2006, p. 54). From this perspective, a mixed methods researcher must contend explicitly with the challenges to internal and external validity, reliability and credibility through sound knowledge of each stage of the

study’s design, implementation and reporting (Cameron, 2011; Long, 2017; Newman, et al., 2013; Onwuegbuzie et al., 2011).

Onwuegbuzie and Johnson (2006) address the noted problems with representation and integration that come from data collected quantitatively and qualitatively, and they present the concept of “legitimation” (p. 55) as an inclusive way to address over-all concerns with validity, credibility, and reliability of mixed methods research. Onwuegbuzie and Johnson (2006) propose nine types of legitimation, and each type provides a nuanced lens that examines the extent to which inferences and findings can be viewed as valid and reliable despite data being collected quantitatively and qualitatively (Onwuegbuzie, et al., 2011). Combined with other frameworks (Cameron, 2011; Newman, et al., 2013), a typology of legitimation has evolved into “a continuous iterative, interactive and dynamic process” (Onwuegbuzie, et al., 2011, p. 1253) of assessing content/construct validity, reliability of findings, and trustworthiness of inferences throughout the cycle of mixed methods research. The following table provides a summary of how each phase of my study met the criteria for legitimation as set forth by Onwuegbuzie and Johnson (2006):

Table 5: Legitimation for Evaluation Feedback Conferences (adapted from Onwuegbuzie & Johnson, 2006)

Legitimation Type-Description	How Addressed in Study
<p>Sample Integration The extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences.</p>	<p>Interview, Survey and Focus Group samples included members of the same population—teachers who have had feedback conferences. The varied approaches yielded consistent inferences.</p>
<p>Inside-Outside The extent to which the researcher accurately presents and appropriately utilizes the insider’s view and the observer’s views for purposes such as description and explanation.</p>	<p>The balance of the emic/etic viewpoints is addressed from the autoethnographic position, further clarified in the Researcher’s Personal Subjectivity Statement. Subjects’ and researcher’s roles vary with each approach.</p>

<p>Weakness Minimization The extent to which the weakness from one approach is compensated by the strengths from the other approach.</p>	<p>Interview weakness/strength: participants' subjective perspectives inherently equivocal/perspectives expose areas of sensitivity for specific elements of conferences</p> <p>Survey weakness/strength: Participants' investment in providing 'accurate' or 'trustworthy' responses/responses reflect objective reporting of experiences</p> <p>Focus Group-Q Sorts weakness/strength: 'constructed' statements may not reflect exact nature of small group of participants' subjective perspectives/narrative quality of 'thinking aloud'—writing thoughts contributes to making stronger inferences than just relying on statistical factors</p>
<p>Sequential The extent to which one has minimized the potential problem wherein the meta-inferences could be affected by reversing the sequence of the quantitative and qualitative phases.</p>	<p>The order of the approaches does not impact the meta-inferences; data analysis shows consistency because it relied on the sequence used: QUAL-QUANT-MIXED</p>
<p>Conversion The extent to which the quantizing or qualitizing yields quality meta-inferences.</p>	<p>The initial quantizing of interview data and literature analysis for the survey instrument reduced the survey items; the qualitizing of the survey items for Q-Sort statements further reduced the items toward interpretable data across all three approaches</p>
<p>Paradigmatic Mixing The extent to which the researcher's epistemological, ontological, axiological, methodological, and rhetorical beliefs that underlie the quantitative and qualitative approaches are successfully (a) combined or (b) blended into a usable package.</p>	<p>Three components were blended to provide a multi-faceted perspective on effective feedback conferences, acknowledging as well as using the paradigmatic assumptions associated with each approach when designing instrumentation and collecting data for making inferences and developing theory.</p>
<p>Commensurability The extent to which the meta-inferences made reflect a mixed worldview based on the cognitive process of Gestalt switching and integration.</p>	<p>The instrumentation and findings for this study relied on the cognitive process of using qualitative data to inform inferences made on quantitative factors and variables, which became a cyclical process that combines inferences for the over-all perspective on the efficacy of feedback conferences</p>

<p>Multiple Validities The extent to which addressing legitimation of the quantitative and qualitative components of the study result from the use of quantitative, qualitative, <i>and</i> mixed validity types, yielding high quality meta-inferences.</p>	<p>Each component of the study went through validity and reliability checks according to the paradigm assumptions: interview data was rendered as the voices of participants (raw data); survey data factors show high Cronbach's Alpha/regression scores; q-sort data factors have medium reliability scores. Separately, each data set stands on own; inferences are stronger when taken as whole</p>
<p>Political The extent to which the consumers of mixed methods research value the meta-inferences stemming from <i>both</i> the quantitative and qualitative components of a study.</p>	<p>The politics of school reform (in the form of APPR law) and teacher evaluation narratives became evident in all three approaches. The voices of the participants (being all teachers) were fore-fronted to give power to those voices about a topic that usually marginalizes the voices and/or are absent from the research literature on the topic.</p>

The one problematic legitimation type for this study would be the Inside-Outside phase and the implications for being the sole coder for portions of the study. I noticed my emic (self-identification as a teacher-participant) perspective needed to be balanced with my etic-researcher perspective when coding and interpreting primarily the qualitative interview data. While I solely coded the interviews, participants in the survey and focus group, in effect, coded the data by virtue of statistical analyses; my role in coding the survey and focus group data focused on labeling the factors based on those statistical analyses. The implications for being the only coder for the purely qualitative interviews relate to how I may not have considered other themes and/or codes which emerged in the interview data set, yet emerged in the survey and focus group data sets.

As the researcher and sole coder for the qualitative data set, I need to address this inside-outside issue with legitimation for possible misrepresentation of the findings that may have impacted how I interpreted what specific narrative data to focus on in my analysis. I would like to note here that I did narrow (i.e. 'scrubbed') some of the interview responses by excluding ad hominem language used by participants when they discussed personal feelings and perceptions

of evaluators apart from reporting on the conference experience in general. My emic-perception of this language, which I shared with my participants, made me protective of my participants as fellow colleagues who obviously were discontented with not only the process, the mitigating factors related to APPR law and how their districts implemented the law, but also extremely discontented with the actual person who evaluated them. My protectiveness extended to how I wanted to be sure the participants' voices would be heard based on the poignancy of their affective response/reporting of their own experiences, and not be 'dismissed' on the basis of 'unprofessional' language targeted specifically toward administrators and/or evaluators. From the etic-researcher perspective, when I reanalyzed the RAP interview narrative data to code the interview data for themes related to the topic and exploration of this study as part of the iterative process of mixed methods research, I focused more on the portions of the narratives that would specifically pertain to objective and subjective reflections, keeping all the language true to the transcripts and fieldnotes. What I found interesting in this step is how my participants self-checked their language in the midst of their interviews by substituting 'professionally acceptable' word usage for what may be considered more raw language (i.e. using "freakin'" in lieu of 'fucking' as an emotional reaction descriptor and/or for emphasis on disbelief), and this self-substitution shows a sense of professional awareness that I had not considered before this re-analysis.

Personal Subjectivity Statement

The topics of teacher evaluation and teacher identity narratives that come from evaluations have been guiding issues for me as both a teacher and a researcher. As a long-standing professional pursuit, I have found engaging with teachers around understanding and being reflective/reflexive about practice to be professionally fulfilling, but also alarming at the

same time. The more I work with teachers on projects like National Board Certification, National Writing Project workshops, and even my home district's teacher-professional development committees, the more I hear the apprehension and feel the tension coming from my colleagues whenever the topic of 'evaluation' or 'APPR' comes up.

During the course of my doctoral program I have engaged a number of teachers about this topic and have noticed that every teacher has a story to tell, some good, some bad, but the tension comes out when the teachers articulate how they think the evaluation is the end-all, be-all of their story, even when they feel the evaluation rating is favorable. During my professional life of 30 years as a teacher, I have always accepted APPR evaluations as being 'part of the job' used by the district to essentially keep track of what teachers do, mostly for audiences other than the teachers themselves; what I have not actually experienced are evaluations that have included me, personally and professionally, in the conversation about how the evaluation itself impacts my practice. I have only experienced limited (if any) input on co-constructing meaning for each 'performance indicator' and 'domains of teaching' on which I have been assessed. I have lost count of the evaluators assigned to appraise my practice, even though I have worked in the same school district for 28 of the 30 years. I have been through enough variations of the process and paperwork involved with APPR that I have an archive of how my district has shifted and changed the process over the years that show how they have used APPRs to meet whatever regulations and/or mandates the State requires. The only component of APPR evaluations that have remained static are the yearly culminations of the process that require a signature saying I accept the appraisal/rating/score that have always named me 'effective.' Whether or not I have challenged or accepted the ratings, I feel there is no space or place in the final renderings of any of my evaluations that allow my teacher's voice to emerge that articulates how I have come to

understand myself through pedagogical decisions and reflection on those decisions. While doing this research and listening to colleagues' reflections, I found that many of us have similar experiences with variations in evaluators' approaches and district evaluation systems, including the 'usual' end result.

As a researcher, I find qualitative approaches align best with how I engage with participants. Teachers are social, and when they trust or confide in other teachers about topics or issues that are important to them, the best way to capture the meaning of those narratives is with social approaches such as interviews, focus groups, document analysis and observation. Even though such approaches open up data interpretation to bias on my part, that bias can be balanced by recognizing, acknowledging and accepting one's own subjectivity in order to avoid letting it skew findings.

Because I am aware of my status as an insider (the emic) and an outsider (the etic) when researching teacher evaluation experiences, I found myself in need of a way to construct a counter-narrative to the predominate power of evaluation narratives, not just for myself but for others as well (Beuving & de Vries, 2015). Because I stand not between the emic and etic but exist in both simultaneously, I am able to approach researching teacher evaluation from the autoethnographic position. Autoethnography is situated within the qualitative research paradigm, and as an approach to research, autoethnography becomes a text that can “democratize the representational sphere of culture by locating the particular experiences of individuals in a tension with dominant expressions of discursive power’ (Neumann, 1996, p. 189)” (as cited in Holman Jones, 2005, p. 765). By doing so, “autoethnography can be defined as a self-narrative that critiques the situatedness of self with others in social contexts” (Spry, 2001, p. 710). A researcher who commits to autoethnographic analysis is committing to laying open the self in

ways that will create “charged moments of clarity, connection, and change” (Holman Jones, 2005, p. 764); “autoethnographic methods recognize the reflections and refractions of multiple selves in contexts that arguably transform the authorial ‘I’ to an existential ‘we’” (Spry, 2001, p. 710-711). An autoethnographic approach, therefore, becomes a means to uncover one teacher story that enhances the telling of other teacher stories, making what happens in teachers’ lives multi-dimensional and generalizable across multiple contexts (Craig, 2007; Kim, 2016; Olsen & Craig, 2001).

In this research project specifically, I explicitly acknowledged my position as a researcher, but also as a teacher with similar-if not the same-concerns about evaluations as many of my teacher colleagues, before conducting interviews and participating in the focus group. By presenting the emic and the etic selves to my participants, I wanted to establish trust so they could feel as though their narratives would be safe with a trusted colleague, not an unknown outsider. During interviews, I opened up about my evaluation experiences (using my emic lens) so that participants would feel a sense of inclusiveness; however, when examining the narrative data from the interviews I consciously used a researcher’s etic lens to uncover the themes that emerged from those narratives, whether or not those themes aligned with my own narrative experiences.

I also participated in the Q-method approach by including a personal q-sort to the focus groups’ collection of sorts. My q-sort was not shared with the groups, but was important to include in the data analysis since my discussion with the focus groups relied on having a professional connection with the group members, not just as ‘the researcher’ (etic) but also as a ‘colleague’ (emic). The decision to include my own q-sort in the focus groups’ analysis opens my interpretations of the sort data to accusations of skewing the data toward my point of view;

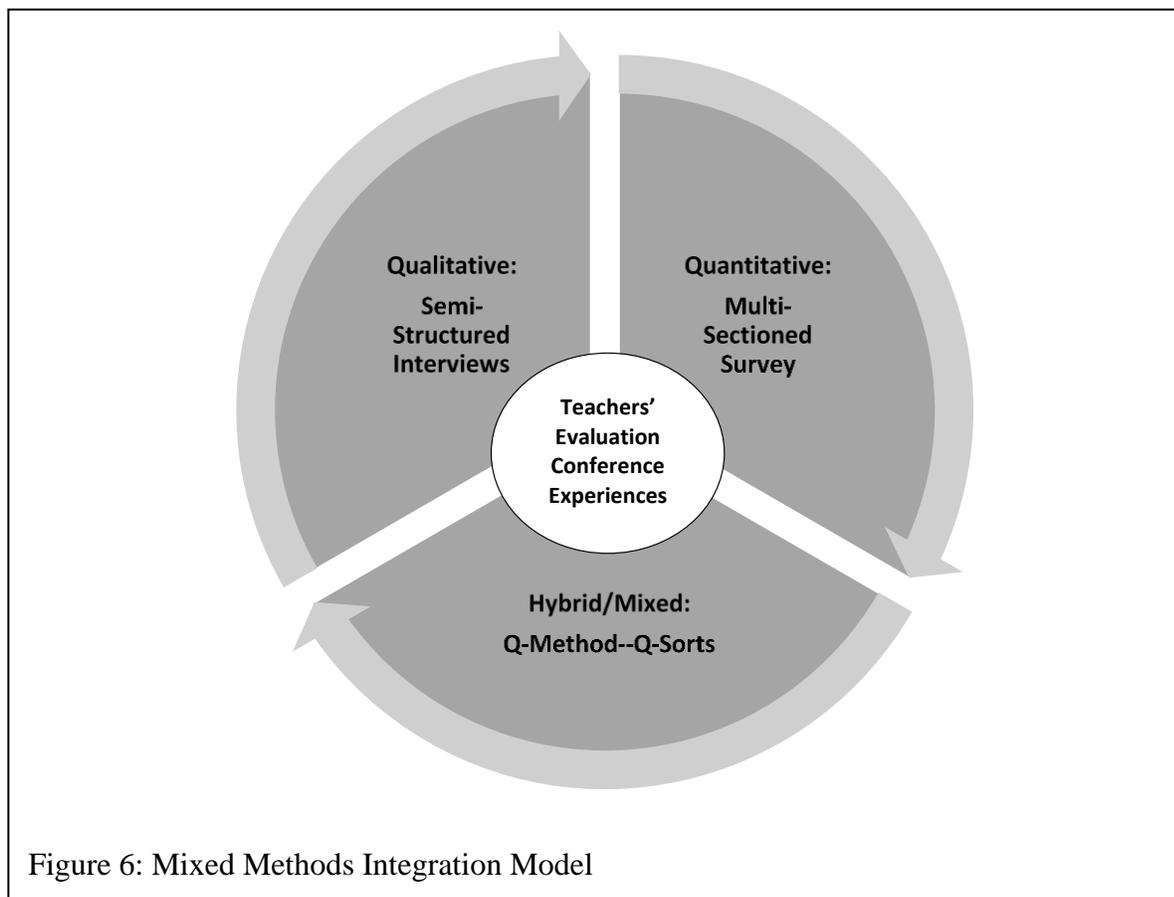
however, I would say that including my own sort is the emic position that is part of the autoethnographic approach. By participating in the q-sort and approaching interviews from an emic position as a like-minded colleague and interpreting all the data from the etic position of researcher, my professional-personal evaluation narrative becomes something to share with other participants as a means to see themselves in the act of revealing their own story in interviews, surveys and focus group discussions. Taking a dual emic-etic position (Beuving & de Vries, 2015; Bogdan & Biklen, 2007; Fine, et al., 2000) allows trust to build between the participants and me, which brought out the groups' subjective thinking over-all and is at the center of taking this autoethnographic approach. After doing this research, I feel committed to helping teachers take back their narratives as a way to uncover an authentic sense of their professional selves (myself included), and use those narratives as a way to counter what they we see as the dominate narratives that currently come from APPR evaluations.

Summary

A mixed methods study is a complex approach that requires diligence for designing, implementing and collecting data that will provide insights on complex phenomenon. A mixed methods research design, therefore, requires the conscious and explicit use of multiple approaches from both qualitative and quantitative research paradigms. The determination of which approach is used first depends on the topic, research problem/questions and purpose for the research study over-all. Once established, the researcher uses the approaches in the order best suited for collecting the data which will fulfill the purpose and answer the research questions. The researcher must also be aware of how the findings are interpreted for representing the perspectives of the participants when reported, and be consciously aware of the impact one's

own subjectivity and biases can have when interpreting the data collected from each of the approaches.

Teacher Evaluation Feedback Conferences are multi-faceted in such a way that no one distinct and separate research approach would suffice to explore the complexities involved. For this reason, this dissertation study consciously uses a pragmatic mixed methods approach to explore teachers' experiences with evaluation feedback conferences to open the 'black box' of how teachers perceive the conference experience over-all. The order of QUAL-QUANT-MIXED approaches allowed for each component of the design to inform the next, working as an iterative cycle, while also allowing the data collected with one approach be validated or disputed by data from one of the other approaches. Taken as a whole, the use of the three approaches gives a holistic perspective on how teachers perceive the evaluation feedback conference and identifies variables and factors that impact how teachers perceive the effectiveness of the conferences for having an impact on their practice.



Chapter 4

Research Findings and Analysis

The design of this study relied on data collected from three distinct and methodologically varied approaches on the same phenomenon: Evaluation Feedback Conferences as part of the teacher evaluation process. Qualitative semi-structured interviews, quantitative survey responses, and focus groups' q-sorts (quantified qualitative item rankings) provided a multi-dimensional rendering of teachers' perceptions of the effectiveness or usefulness of evaluation feedback conferences. The decision to use a mixed methods research design was based on how each approach had the potential to reveal common themes of what happens in feedback conferences, and at the same time provide space for any nuances on those common themes which monomethod research approaches were less likely to uncover. The following sections of this chapter examine data from each approach and an analysis of the data. The summary will examine the most significant commonalities as well as any unique perspectives that reveal the complexity of teachers' sense-making of their experiences with evaluation feedback conferences.

Semi-Structured Interviews

The first data set collected was semi-structured interviews from my Research Apprenticeship Project (RAP) focusing on teachers' perceptions of the evaluation conference and whether/how that conferencing experience impacted their practice. The interview questions were developed to specifically focus on how teachers felt about those conference feedback conversations in relation to New York's Education Law (APPR) mandates for summative performance ratings. The collected data consisted of short, anecdotal narratives about APPR feedback experiences, focusing on how the teachers felt about the observation-evaluation

process, their participation in the process, and the way lead evaluators conducted the feedback conference (see Appendix A: RAP Interview Guide).

In total, eight interviews provided data that went through multiple coding, categorizing and theming cycles (Saldaña, 2013). Interview transcription data were organized separately as codebooks for each participant to gauge how each participant was aligning with emerging objective and subjective themes. I went through the narratives initially, as part of the RAP, for themes that described the over-all evaluation experience and then a second time for how the narratives addressed more self-reflective, personal themes. For the next cycle of coding for the RAP, I adjusted the categories to more explicitly address the RAP research questions, paying specific attention to how the participants revealed issues with evaluations generally, not just within the school year during which they were interviewed. These codes and themes became the starting point for my dissertation study because of the way the participants opened up about how they felt and struggled, personally and professionally, with the way the APPR evaluation process was being implemented in their districts. Even though I did not have a large amount of data from the interviews, the complexity of their experiences came through the narratives enough to show that the nature of teachers' feelings about evaluation feedback conferences, how evaluators approach conducting those conferences, and the evaluation process in general warrant further research.

As the first phase of my dissertation study, I re-analyzed/coded the RAP interview data, using the dissertation questions as an interpretive lens, for participants' reactions/responses to their evaluators' approaches to feedback conferences as a way of implementing the APPR process, with specific attention to discourse that reflected specific affective responses to evaluators' approaches. Analyzing the narrative data from this position, I noticed how the

participants were expressing what they felt personally and affectively about their experiences, even though they did not always explicitly name what they were feeling or what their affective responses represented, unless asked directly. The themes I noticed from reanalyzing the interview data focus on contextual issues that prompted changes to the APPR process, authenticity of evaluators' appraisals as part of the process, issues with trust-power/control over individual evaluations based on how evaluators implement the process, impact of evaluators' consistency for implementing process and/or protocols, and perspectives of self-worth/efficacy reflected in evaluators' assessments of practice.

Contextual Issues

Contextual issues with changes in the APPR process across the participants' districts emerged as a minor but noticeable theme specifically prompted by the interview questions. Each participant noted a change in the conducting and/or purpose of the evaluation conference from previous experiences, and cited the APPR regulations, some explicitly and some implicitly, as an underlying reason. One participant explicitly commented on how "since the new APPR system came into play, I have not received one piece of suggestion, one piece of advice, one, 'I noticed you did this. Why would you do that?' No question about my practice." Another participant pointed out that, "I personally find the whole thing incredibly ironic because it has been my understanding that the new APPR is designed to be more objective than our old system of evaluation based one hundred percent on an observation" and yet the 'new' way still retained the subjective observation for determining sixty percent of the final efficacy score¹. The narrative comments seem to be addressing the contextual issues with how the APPR requirements are

¹ The version of the APPR law the RAP interview participants refer to is 3012-c/d, which used a formula of 60% observation of practice with 40% student growth score to determine the over-all HEDI effectiveness/summative rating

shifting, yet evaluators' approaches to the conference as part of the APPR process still seem to favor summative over the formative purposes, which may be contributing to the tension felt when dual formative and summative purposes are being addressed simultaneously to fulfill APPR requirements.

Authenticity, Trust=Power/Control

Another theme I noted in the narratives was participants' feelings towards evaluators' authenticity when evaluators engaged them in discussions about their individual practice. I used the code of *authenticity* (see "Definitions" in Chapter 1) for this theme since I noticed how participants articulated strong affective responses about their experiences when they felt evaluators were not being genuine or sincere about the appraisal of their practice. For the whole process, or even a piece of the process, to be considered 'authentic' means the evaluator puts effort into understanding what teachers do and is able to give sincere, genuine, and mindful feedback that shows awareness of what the teacher does in day-to-day practice. The interview participants do not explicitly use the term "authentic" or "authenticity" when reflecting on their experiences or give their impressions, but I interpreted their critical language and predominately negative tone in the following excerpts as representative instances from the data set that imply feelings about evaluators' authenticity (as I have defined it) when giving feedback to the participants:

Interview #1: As far as me learning from him—nothing. Because I feel he has nothing to offer me. Again, it's not because I'm being snotty or narrow-minded, I admit my flaws, I just don't feel like it's a true evaluation because he has standards he says to every single person that goes in there. He throws out a couple of catch phrases and it's a shame. [It] becomes not only a waste of time [but] almost detrimental because he doesn't listen to a word I say. He doesn't have a connection with me what-so-ever.

The notable critical language, “true evaluation,” “it’s a shame,” “waste of time,” and “detrimental” indicate to me that the evaluator was not sincere or genuine when giving feedback.

Interview #7: I do remember another assistant principal evaluating me and I felt like when they wanted to give me feedback, they were looking for something... ‘what negative thing can I say so I’ve got to put something down.’ And I remember her saying...something but it was so crazy stupid, like really, that’s your feedback? That’s your suggestion for me? I don’t think that she brought anything to the table. She was very young. I think she was only a teacher for three and a half years and then she became an assistant principal. She didn’t have anything to offer, maybe because I was a lot older than her that I think she really just wanted to really give me something...something to say, because she had to.

In this excerpt, I interpret the participant’s meta-reflection, “ ‘what negative thing’ ” as conveying a critical position and negative tone toward what the evaluator said, while also noting that the evaluator’s feedback was “something but it was so crazy stupid” and “didn’t have anything to offer,” indicating to me that this participant felt whatever the evaluator had to offer as feedback lacked sincerity. What the two excerpts represent are responses to experiences during which evaluators showed no mindfulness, thus no authenticity, so the participants dismiss the feedback given during the conference as useless.

When the participants feel the authenticity of the evaluation/conference is in question, they make comments on how trust in evaluators, and collaboration on the evaluation process over-all, affect how they feel:

Interview #1: He and I have a decent relationship. I’ll show him lessons that he didn’t observe, and things I can do and I don’t lie, but I could and he wouldn’t know it. In theory, he’ll say ‘okay’ and check it off but I don’t feel 100% confident that he’s checking that so that is why I bring copies. I have kind of just lost faith.

This participant expresses how she feels a level of trust in the evaluator based on the “decent relationship,” but there isn’t enough sincerity or genuine appraisal of practice (authenticity) that would make her feel “confident.” The most affective response I note with this comment is how

the participant has “just lost faith” in the teacher-evaluator relationship that, in turn, has an impact on what level of trust she feels.

Interview # 2: This is my life. And for her to be messing around with it, for whatever her little power trip is, I don't know what it is, but for whatever it is, with everything that's going on now, it's even ten times worse. No, she's very black and white cut dry.

This participant makes an explicit connection between “life” and what evaluations say about her “life” based on what the evaluator does. As expressed, I interpret the very critical and negative tone of this excerpt about the evaluator's behavior as showing no trust in the evaluator over-all, especially in the way the teacher perceives the evaluator being on a “power trip” and “very black and white cut dry” in the appraisal.

Interview #3: I have a voice, whether or not they're going to make any changes based on the conversation. I've had an administrator, after a conversation, be willing to move [a score] but I've also had times where, ‘No, this is what I saw and this is my interpretation of this indicator and this is what the score is.’

This participant does recognize her voice in the evaluation, yet notes there is no way to know if the evaluator will be willing to collaborate or “make any changes based on the conversation” that will show a sincere understanding of what she brings to the conversation. Without having the trust in the evaluator for understanding or listening to the teacher's side of the evaluation, there cannot be any way for the evaluation rating to be authentic; it is insincere and ingenuine to marginalize a teacher's voice to pronounce “what the score is” without considering what the teacher brings to the evaluation conference discussion.

The theme of trust is connected with issues of control and power over, and purpose for, the evaluation process. Issues of trust, power and control come out in how the participants react to the approach evaluators use to determine ratings, which lead to interpersonal/contextual issues that impact how the participants accepted the over-all ratings:

Interview #1: And then you had to prove him wrong. So he had that kind of approach. You had to kind of bring proof which I always took personal offense to because it's like not only do I have to do my job and do it well and then I have to go and prove to you that I do it well because you can't figure that out.

The participant's critical tone comes out when she notes taking "personal offense" to the evaluator's approach for collecting evidence of practice that must "prove" the efficacy of her practice.

Interview #2: Because I need to do what's right for them [students] and not what somebody in some stupid office is thinking might be right to help themselves. Because they're all out for themselves, every one of them now; they have to be. And so they want to try to control the process. But you can't without our input because we're the ones doing the work with the kids. And they don't want to listen to our input.

In a much more explicit way, this participant voices a negative assessment of her evaluators' intentions for the APPR process, especially when she notes how, "they're all out for themselves, every one of them now" which connects the contextual changes in the APPR process with the way evaluators are conducting reviews. This participant's (Interview #2) affective responses throughout the interview could be described as the most critical and negative of the group, but also representative of a common perception for having to "prove" professional judgment in over-all practice to evaluators when the summative appraisal of their practice does not include their input. As a group, the predominate feeling of trust, while not actually mentioned explicitly as 'trust,' was connected to underlying conditions of control over the conference and evaluation situation by evaluators who did not acknowledge the teachers for knowing what is best for students and their teaching practice. This perceived lack of understanding teachers' professional knowledge (which is an actual domain of teaching on all the approved APPR rubrics) by the evaluator(s) directly impacted how the interview participants took ownership of the evaluation context and/or process in general.

Consistency of Implementation

Another major theme that came through the interview data is participants' feelings about the consistency (or inconsistency) of evaluation methods and protocols for gathering evidence of practice. When teachers perceive the actions of evaluators to be subjective, they voice their frustrations with having to 'figure out' what to expect from the evaluator because there is a lack of consistency in the process:

Interview #2: Because my experience has been, they may read the rubrics in a similar way and use some of the same sort of buzz terms, but there's certain strategies that one administrator might have liked, might have used as a teacher and thought, 'well these were effective for me and I'd like my staff to use this.' So you get used to that administrator's way and then you get a new administrator who comes in and you've gotten comfortable with teaching in a way that using strategies of your old administrator. And now a new one comes in and now they're talking about seeing these things in a classroom, 'I'd like to see teachers do this; I'd like to see teachers do that.' And you're like, wow, I just started implementing all these things that the old administrator liked and now I've got to go back and do some of these things instead. So, I guess it's a little bit of a game sometimes that you're playing.

This participant's over-all response to the lack of consistency between evaluators' approaches to appraising specific teaching strategies echoed through all the interview data. The inconsistency for protocols used and the variations on what constitutes effective teaching strategies contribute to the frustrations felt by all the participants. For this participant, and the one following, the subjective nature of how evaluators conduct evaluations over-all is connected to the inconsistencies in applying evaluation processes and/or protocols which impact how teachers respond to the appraisal of their practice:

Interview #8: Given that I have now had three evaluators under this more objective system, I can say that in my personal experience, it is still incredibly subjective. And regardless of how it's structured, or what rubric is used, or how the evaluators are trained, I think that evaluating a person's teaching is always going to be sort of subjective.

These two comments highlight how teachers react to the apparent subjectivity of the evaluators, where it feels “like a game” with each different evaluator that results in an inconsistency that teachers cannot control. The consistency/inconsistency of evaluators’ approaches across and between districts impacts teachers’ perceptions of the evaluation process, as a whole, for not being able to give an authentic (i.e. sincere and genuine) appraisal of their performance. When teachers perceive inconsistencies within and between systems, then the authenticity and efficacy of the evaluative process becomes undermined by the continuously shifting nature of the how that evaluation is conducted.

Teachers’ Self-Worth and Efficacy

What also came through the narrative data were how participants feel a tension between the evaluators’ rating of performance and their self-perceptions of efficacy not reflected in those ratings. I notice this tension comes out when the participants articulate how the evaluators’ feedback would seem disconnected from their day-to-day teaching practice because evaluators are focused on fulfilling their ‘check list’ items of the APPR process (language noted in **bold**):

Interview #2: Last year, learning targets were the focus, [and] **we had no input**. It is part of the problem. And so, that’s why it feels like **we never have input on anything**.

Interview #4: I almost felt like she couldn’t like really appreciate the amount of work they [students] put into it because **she was just focused on how many were in position**.

Interview #5: Those administrators **never came in the room**. The only thing they’re basing it off of was that **one time**.

Interview #6: And they’re just typing, typing, typing, typing, or writing, writing, writing, writing. **Gone** are the days when you can go, ‘Ok, the principal is coming, if I do a lesson, he’s going to put the pen down and get up and go chat with the kids.’

Interview #8: I just can’t describe to you the **number of boxes that I filled in** and, more importantly, that I kind of approached it at some point as filling in boxes because it was an **excessive amount of paperwork**. **It did not feel like it**

was centered on my improvement or anything that I could then bring with me to the classroom; it was really more just a task.

The participants' reactions to feedback reveal a conflict between what they feel they do on a day-to-day basis and their perceptions of the evaluators' lack of understanding their day-to-day practice. In some instances, participants' critical tone showed how they felt they had no reason to accept what the evaluator said since the ratings showed no understanding of the teachers' practice. When participants note the evaluators' behaviors show an inconsistent application of teaching rubrics that lack equity/fairness or use arbitrary protocols for observations that show a lack of awareness of a teacher's day-to-day practice, the teachers' perspectives of the evaluation process become damaged to the point anything having to do with evaluations is considered suspect.

The themes which emerged from the RAP interviews, thus, provided a framework for extending and furthering the research on the effectiveness and usefulness of the feedback conference from the *teachers'* perspectives. Taken as shared perspectives on a common experience, the narratives clustered around themes of specific feelings towards how evaluators conducted feedback conferences (especially interpersonal feelings towards specific evaluators and trust) and mandated requirements for APPR ratings related to the culture and context of their individual teaching situations. Those subjective narrative themes gave voice to teachers' feelings that are absent in the literature for understanding how the feedback conference can address issues with conducting APPR evaluations over-all. Even though eight narratives are not enough to generalize across an entire population (i.e. teachers in NYS who are evaluated with 3012-d), the themes which did emerge shed light on issues and tensions that I felt are worthy of further research.

Evaluation Feedback Conference Survey

The next set of collected data consisted of teachers' responses to a multi-sectioned survey (see Appendix C: Teacher Evaluation Conference Survey, 2019). Over-all, 39 survey responses were compiled as an excel file from Qualtrics and then entered into SPSS for statistical analysis. Prior to performing a statistical analysis of the data, I assigned codes and categories for each section of the survey, in much the same way as I coded interview data (Saldaña, 2013), in anticipation of over-arching themes that would address the research questions, and then the category-codes were assigned to the sections of items after the data was entered into SPSS. The codes identified groups of items that specifically address the evaluators' approaches for addressing performance, engaging teachers in the evaluation process, and using interpersonal/relationship building strategies in the conference process. Other codes identified items that focused on what topics the teachers and evaluators had discussed in the conference and various general aspects of the teacher's relationship with his/her evaluator and his/her fellow teachers. The last set of codes identified the subjective reflections on the effectiveness of the conference as part of the evaluation experience. This coding process became the outline for trying out combinations of items as variables in the statistical factor analysis process.

The following sets of tables below display the raw data of all the collected responses, complete and partial, that show how participants responded to each item. The tables are organized by groups of survey items that *eventually* factored into clusters within and between sections as a result of factor analysis. Tables A—I provide raw data on the factor clusters around evaluator's behaviors and approaches to conducting conferences; tables J—L provide raw data on three factor clusters representing conference effectiveness measures. Factor labels are bulleted, and the items under the factor labels provide the context of that particular factor.

Within each table, the items are listed by mean scores of the responses in that factor cluster, beginning with the item rated with the most accurate, extent, and effectiveness within the respective survey sections. Mean data scores with an asterisk indicate the “Do Not Know/Cannot recall” or blank responses to an item are not being included in the calculation of the mean for that item. Tables without the “Do Not Know/Blank” column indicates all participants (from all complete or partial surveys) responded to the items by selecting a rating for the items. Each table provides the “Percent responses.” representing the rating scales for the sections associated with those items. For clarity purposes, any of the “percent” totals may equal 99 or 101 because of rounding purposes. The standard deviation by item within that cluster follows the “Percent Responses” in each table, with the discussion of the raw data following each table as an initial analysis of the data related to the eventual factor clusters.

The first two sections of items asked participants to report on approaches used by their evaluators when conducting evaluation conferences. For these sections, items address observable and/or reportable approaches which educational leadership literature espouses as sound practices for evaluators to use when conducting conferences. The following Tables 6A and 6B display items that factored on general approaches evaluators sometimes use when conducting conferences.

Table 6A

Raw Data: Percent Survey Responses to Items for Independent Factor "Identifies and Addresses Evaluator's Concerns" (n=39)

Items	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
Discussed aspects of teacher's teaching that [the evaluator] considered to be areas of concern	*3.13	11	11		24	18	37	1.37
Discussed what the teacher would do to address areas of concern.	*3.17	8	13		16	18	39	1.46
Discussed concrete examples of anything [the evaluator] considered to be an area of concern	*3.25	16	11		16	21	37	1.48
Discussed which concerns or steps for improvement should receive highest priority	*3.43	8	13		16	18	45	1.37

The mean responses to the items in Table 6A indicate that evaluators, more often than not, do not seem to make their issues and concerns the focus of the discussion. The responses to the items in Table 6B indicate that evaluators place more emphasis on the teacher's accomplishments and strengths:

Table 6B

Raw Data: Percent Survey Responses to Items for Independent Factor "Evaluator Identifies and Addresses Teacher's Accomplishments/Strengths" (n=39)

Items	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
Expressed appreciation for the work the teacher does	*1.32	79	16	3	3		3	0.78
Discussed the teacher's general accomplishments and strengths	*1.43	66	26	5	3		3	0.83
Discussed concrete examples of the teacher's accomplishments or strengths	*1.70	47	42	5	3	3	3	0.90
Discussed ways of building on the teacher's accomplishments and strengths	*2.36	29	32		18	8	13	1.35

The standard deviations for the set of items in Table 6B are the most notable of the over-all factor clusters because of the general consensus across the majority of participants on this set of evaluator behaviors. This consensus on this group of items seems to imply evaluators are putting more emphasis on teachers' accomplishments and strengths than on their own concerns when conducting feedback conferences.

The survey's second section asks participants about the approaches taken by evaluators to involve them in the conference. The items in this section focus on discrete actions, discussion protocols, and dialogue/constructive conversation strategies that educational leadership literature

urges evaluators to purposefully employ. Tables 6C and 6D, below, display items that factored on general approaches that evaluators sometimes do to involve the person being evaluated:

Table 6C

Raw Data: Percent Survey Responses to Items for Independent Factor "Attention to Teacher's Input" (n=39)

Items	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
Paid close attention to what the teacher had to say	1.66	61	24	8	5	3		1.02
Used strategies like paraphrasing, maintaining eye contact and other non-verbal cues to convey attention to the teacher's opinions	*1.68	66	16	8	5	5		1.17
Was open to teacher's opinions, even if they might differ from her/his own	1.79	58	21	13		8		1.19
Used open-ended questions that invited discussion rather than assertions that would close off discussion	1.82	55	21	16	3	5		1.14

The mean for the items in Table 6C indicates that most evaluators seek to engage teachers by paying attention and using approaches that would open the discussion to what teachers wanted to talk about. Participants notice evaluators are listening and report that evaluators employ

discussion strategies or approaches that should be used when paying close attention to teachers'

input:

Table 6D

Raw Data: Percent Survey Responses to Items for Independent Factor "Seeking Teacher's Input" (n=39)

Items	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
Encouraged the teacher to provide information relevant to the topics being discussed	2.39	32	24	29	5	11		1.29
Asked the teacher before the formal observation what the evaluator should look for	2.45	37	21	16	13	13		1.45
Encouraged the teacher to suggest options for addressing identified concerns	*2.47	26	26	16	11	11	11	1.35
Asked teacher before the conference to be prepared to discuss anything the teacher thought needed attention	2.53	26	32	21	5	16		1.37
Urged the teacher during the conference to identify goals or concerns the teacher wanted to discuss	2.58	26	26	24	11	13		1.35

Urged the teacher to identify anything that might affect the teacher's ability to teach effectively	2.89	26	16	24	11	24		1.52
Invited the teacher to choose among different options for addressing identified concerns	*2.97	11	21	29	8	16	16	1.28
Emphasized the need for the teacher and evaluator to reach conclusions both could support	*3.00	13	13	34	8	16	16	1.30

The mean scores for this group of items, spread across 'very accurate' to 'as accurate as not,' seem to imply that most evaluators are using noticeable 'personal' approaches when conducting conferences, especially with 'paying attention' to what teachers were saying, more so than trying to get input from them during the conference. However, even though evaluators are using these approaches there is no way to know if participants actually feel they are being engaged in the discussion when/if evaluators use those particular approaches.

One survey section asks participants to report on the extent evaluators would focus on particular topics during the conference. The topics represent a broad range of observable actions and teaching rubric domains addressing student learning objectives and processes, teaching practice, and professional responsibilities. The following Table 6E displays items that factored on the extent to which evaluators focused on topics related to teaching practice:

Table 6E

Raw Data: Percent Survey Responses to Items for Independent Factor "Discussion of Teacher's Practice" (n=39)

Items [While conducting the conference, the evaluator...]	<i>M</i>	Quite a lot	A fair amount	To some extent	Not much	Not at all	Do not know/ blank	<i>SD</i>
Discussed whether and how teacher's students actively engaged in learning	1.53	58	32	11				0.69
Discussed other data or information, besides test scores, on what students have learned	*1.58	26	37				37	1.28
Discussed how the teacher assesses students' learning	1.68	47	37	16				0.74
Discussed teacher's classroom management and/or relationships with students	1.71	50	29	21				0.80
Discussed how the teacher adapts instruction for different students	1.87	45	32	16		8		0.96
Discussed teacher expectations for students	2.00	40	37	13	5	5		1.12
Discussed whether students are meeting expected learning standards or objectives	*2.84	13	21	42	11	11	3	1.14
Discussed teacher's knowledge of the content taught	2.53	24	37	13	16	11		1.31
Discussed how the teacher plans lessons	3.53	5	21	21	21	32		1.28

The responses show a distinction between items evaluators addressed in the conference that are more observable (and do not require teacher input for the discussion) and other items that are less observable (and more difficult to evaluate without teacher input). The mean scores show this

distinction with the most significant topics evaluators address focusing on student engagement/assessment and classroom management/differentiation of instruction (the first three items that are more observable). The next significant topic (the fourth item) focuses on student output/productivity, which needs more input from the teachers for assessing. The topics related to specific domains of teaching (the bottom five items) are more subjective on the part of the evaluator and need more input from the teachers; without teacher input, the evaluation of those items become more difficult if only observation evidence is used to determine ratings for those items. The mean scores seem to suggest that evaluators are focusing discussions more on what they observe teachers doing with and for students than on what teachers can say and reflect on for what they are doing with and for students (including the lesson planning and other student data used for instructional planning).

One item from this section of the survey (i.e. topics of discussion), ‘discussed student test scores,’ did not factor with any of these items, nor did it factor with other items. However, this outlier item did relate to other outlier items from sections on evaluators’ general approaches. Even when these outliers were combined as a composite factor, the items did not load as a factor or show reliability as a composite scale when factored together:

- GENERAL APPROACH:
Let the teacher know before the conference anything evaluator wanted to discuss
Explained what the evaluator wanted to accomplish in the conference
- GENERAL APPROACH: Discussed limited number of rubric elements the district uses
- GENERAL APPROACH: Discussed some/all of the ratings the evaluator planned on giving
- TOPIC: Discussed student test scores

The data analysis for these collective items indicates participants acknowledge that some of the approaches and topics that address compliance with APPR mandates are a part of the conferences, but the data do not differentiate if those approaches and topics are more for the sake

of the evaluator fulfilling APPR requirements or for supporting teachers' improvement in practice. For this reason, I decided to leave these items out of the over-all interpretation of the data.

There are other outlier items from the general approaches and general topics sections which are notable for how they address more summative purposes for evaluation conferences as related to the APPR requirements and became a factor cluster. As a factor cluster, the four items reflect participants' experiences with evaluators who approach the conference as a way to discuss performance ratings. The following Table 6F displays items that factored across survey sections on approaches and topics:

Table 6F

<i>Raw Data: Percent Survey Responses to Items for Independent Factor "Summative Ratings/ Assessments" (n=39)</i>								
Items	<i>M</i>	Quite a lot	A fair amount	To some extent	Not much	Not at all	Do not know/ blank	<i>SD</i>
Explained reasoning for the ratings evaluator planned on giving	*2.15	32	37		11	8	13	1.31
Discussed all/most of the evaluation rubric elements the district uses	*2.44	24	39		16	11	11	1.27
Discussed the ratings the evaluator assigns	*2.62	18	34	21	13	11	3	1.26
Discussed compliance with district/school policies	3.45	11	21	13	24	32		1.41

Teachers are noting that evaluators use the conference for summative ratings discussion, but there is no clear indication how evaluators are incorporating this ratings discussion with the other topics. The discussions seem to be focused mostly on rubric elements and evaluators' rationales for giving the ratings according to the APPR rubric used by the district. As an approach, these

discussions seem to over-take discussions about the assigned ratings and how teachers comply with policies.

Two sections of the survey ask participants about the general interpersonal approaches used by evaluators and general situations/relationships between evaluators and teachers that may impact the evaluation conference over-all. Compared to other sections that asked for objective reports of observable behaviors, these items were more subjective, asking participants to judge the behaviors from a position of fairness and equity based on approaches and relationships that are developed within the context and culture of the participants' teaching situation. The following Tables 6G and 6H display items that factored on general approaches the evaluator took while conducting conference and general situations and relationships that affect conducting conferences:

Table 6G

<i>Raw Data: Percent Survey Responses to Items for Independent Factor "Evaluator's Fairness" (n=39)</i>								
Items	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
The evaluator was respectful in how he/she shared opinions	*1.30	74	21			3	3	0.62
The evaluator showed that she/he had the interests of students in mind	*1.32	74	18	3	3		3	0.67
The evaluator used accurate information	*1.38	66	29		3		3	0.64

when discussing performance								
The evaluator appraised performance fairly	*1.59	50	42	3		3	3	0.80
The evaluator showed that he/she is a good judge of teachers' effectiveness	*1.65	53	32	11		3	3	0.89
The evaluator offered useful perspectives on things she/he observed in teaching	*1.89	47	26	16	3	5	3	1.13
The evaluator took the time to give each teacher's evaluation careful attention	*2.00	37	37	16	3	5	3	1.08
The evaluator helped see situations teachers face from different vantage point	*2.17	29	34	24	3	5	5	1.08

The highest consensus among the participants about specific approaches being used, as noted by the standard deviation, relate to how evaluators showed respect, interest in the students, and accuracy about what was discussed. These over-all responses indicate most participants feel evaluators are equitable and fair, which match with the responses that reflect how participants feel the evaluators are aware of what they do in practice:

Table 6H

Raw Data: Percent Survey Responses to Items for Independent Factor "Evaluator's Awareness of Teacher's Practice" (n=39)

Items	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
showed an awareness of what teachers have done to improve teaching	*1.74	50	24	13	3	3	8	1.01
understood the instructional challenges teachers face	1.92	47	26	18	3	5		1.12
understood the curriculum teachers are responsible for following	2.21	34	32	18	11	5		1.19
conveyed a clear vision of what he/she wants students and school to accomplish	2.45	24	34	24	11	8		1.20
asked for teachers' advice on issues that affect them	2.45	24	34	21	16	5		1.18
observed teaching on a regular basis	2.87	18	26	16	29	11		1.32

This second set of items in Table 6H show evaluators are generally paying attention to the teachers, and teachers notice this when in the conference. However, the data also indicate that there is an inconsistency across specific behaviors and interpersonal approaches that show

awareness of teachers' practice when evaluators do not observe teachers on a regular basis or ask for teachers' advice on issues that affect them.

A cluster of related outlier items from across sections reflect how evaluators approach topics that show support for teachers' improvement through on-going collaborative professional development between teachers and engagement with evaluators on plans for improvement. The following Table 8I displays items that factored together on approaches, topics, situations, and relationships:

Table 6I

<i>Raw Data: Percent Survey Responses to Items for Independent Factor "Provides for Follow-up"</i> (<i>n</i> =39)								
Items [While conducting the conference, the evaluator...]	<i>M</i>	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Don't know or blank	<i>SD</i>
encouraged teachers to turn to each other for advice	2.16	32	32	29	5	3		1.03
encouraged teachers to try new ideas, even if doing so might mean making mistakes	*2.32	32	26	21	13	5	3	1.23
made useful suggestions for specific things teachers might do to change teaching	*2.38	24	32	29	8	5	3	1.11
encouraged teacher to work with other teachers	*2.78	8	16		18	18	39	1.35

to address concerns teachers share								
made sure to agree on what steps the teacher would take as a result of the conference	*2.93	18	18		18	16	29	1.48
discussed specific things the evaluator would do to support the teacher's effort to improve	*3.00	11	18		13	13	45	1.48
discussed teacher's plans for improving teaching	3.34	13	15	26	21	26		1.36

The participants note how evaluators are having discussions about improving practice in an encouraging way; however, the responses also indicate that specific steps evaluators would take to support that improvement or even elicit teachers to make a plan for themselves to improve is happening less often. What is not indicated by the responses is why the evaluators would approach the conference in such encouraging ways and then leave out the formative discussions for how to improve practice.

The last section of the survey asks teachers about 'how effective' the feedback conferences were for promoting particular outcomes. The items for this section of the survey asked participants for their subjective assessment of the conference experience based on effectiveness 'criteria.' The following Tables 6J, 6K, and 6L display items that factored on how effective teachers thought their most recent evaluation conference was:

Table 6J

*Raw Data: Percent Survey Responses to Items for Dependent Factor
“Effectiveness=Strengthening Teaching” (n=39)*

Items	<i>M</i>	Very Effective	Fairly Effective	As effective as not	Fairly ineffective	Not effective at all	blank	<i>SD</i>
Providing an opportunity for serious discussions of different approaches to teaching/ other work with students	2.50	24	29	29	11	8		1.20
Helping teachers develop own solutions for addressing identified goals or concerns	2.74	18	29	29	8	16		1.31
Helping the teacher to improve knowledge and skills	2.76	18	24	34	11	13		1.26
Helping the teacher make sense of problems/ concerns teachers face in teaching/ other work with students	2.89	13	29	29	13	16		1.27

The responses to the items in Table 6J indicate that participants feel evaluation conferences do, indeed, support their work towards improving practice. Table 6K, below, displays items for the second effectiveness measure and imply how the conference is effective when it provides the opportunity for building and cultivating evaluator-teacher relationships. As another set of items on measures of effectiveness, the mean scores to these items indicate the participants recognize the opportunity to develop the professional relationship with evaluators during the feedback conference; these interpersonal relationships can lead to trust and professional reflection on practice. However, the responses also imply that conferences are not necessarily effective for collaborative professional development.

Table 6K

Raw Data: Percent Survey Responses to Items for Dependent Factor "Effectiveness=Cultivating Teacher-Evaluator Relationships" (n=39)

Items	<i>M</i>	Very Effective	Fairly Effective	As effective as not	Fairly ineffective	Not effective at all	blank	<i>SD</i>
Providing recognition for the teacher's efforts	1.82	58	16	16	8	3		1.14
Providing the teacher with an opportunity to reflect on own performance	1.87	45	32	16	8			0.96
Fostering trust between teachers and administrators	2.05	42	21	32		5		1.11

Providing teacher with feedback that could be used to strengthen teaching/ other work with students	*2.27	32	21	34	8	3	3	1.10
Identifying ways to further the teacher's professional development	2.95	16	21	32	16	16		1.29
Arranging for the teacher to get help or resources to improve teaching/ other work with students	*3.11	5	29	32	13	18	3	1.20

The third effectiveness measure items, displayed in Table 6L below, focus on the summative purpose of evaluations over-all:

Table 6L

Raw Data: Percent Survey Responses to Items for Dependent Factor "Effectiveness =Accountability" (n=39)

Items	<i>M</i>	Very Effective	Fairly Effective	As effective as not	Fairly ineffective	Not effective at all	blank	<i>SD</i>
Holding teachers accountable for students' learning	2.08	5	26	40	29			1.01
Holding teachers accountable for the teaching strategies used	2.32	5	24	34	34	3		1.04

As an effectiveness measure, participants are noting accountability for student learning and teaching strategies as a necessary part of the evaluation process, and the evaluation feedback conference presents an opportunity to fulfill that part of the process. However, the responses do not indicate if the participants feel accountability comes from how evaluators approach the conference or from their investment in the process.

The survey items on conference 'effectiveness' were purposefully placed at the end of the survey, and were intended to reflect more subjective statements on the purpose and function of evaluation conferences within the over-all evaluation process. Based on how the items focus on the 'criteria' of effectiveness, the responses seem to imply teachers recognize the conference as an experience that contributes to cultivating an effective relationship between them and evaluators when evaluators use approaches that acknowledge their work. The responses also imply teachers recognize the conference as providing an opportunity to strengthen their teaching. The more complex preconception of conference effectiveness relates to how participants report that the experience is not as effective for helping them with professional development and access to resources that would contribute to improving practice, even though they report conferences are effective with relationships that strengthen teaching.

As noted earlier, the raw data are organized according to the extracted factors. These factors are based on how items loaded within and between survey sections using SPSS data analysis software program. I specifically used the structure matrices from the PAF analysis to determine which items to include for the strongest factors. From the structure matrices, composite scales were computed by adding participant scores (replacing missing data with the mean score) and then dividing by the number of items in the scale. Since there are multiple items for each factor, the labels given to the factor are meant to represent the broad, underlying theme

(i.e. coding) of the items which cluster on that factor, using the highest loading item of the cluster as the focusing thematic concept.

Once composite scales were determined, I used Cronbach's Alpha to check for the highest reliability for the individual scales. If the Cronbach's Alpha showed an increase in reliability with an item deleted, then I performed re-calculations of the composite scales with items deleted. I reverted to the initial composite scale if there was a decrease in reliability, no significant change to the reliability, or significant structural change in the composition of the scale when deleting an item (i.e. too few items in scale with deleting ones indicated by Cronbach's Alpha). After determining the composite scales and checking reliability, twelve factors were extracted from the data representing nine independent factor variables and three dependent factor variables.

The factors extracted from the data, from four of the five sections of the survey, give the independent variables for the study. These independent variables coalesce around how evaluators approach conducting the conference, involve the teachers in the evaluation process/conference, include specific topics as protocol for/during the evaluation process, and develop relationships with teachers based on the reported culture and context of districts that impacts evaluation conferences/process:

Table 7: Summary of Independent Factor Variables

Factors: Independent Variables	Variable Mean	Variable Standard Deviation	Cronbach's Alpha (Reliability)
Evaluator's Fair Appraisal	1.691	0.763	.941
Attention to Teacher's Input	1.734	1.025	.929

Addresses Teacher's Accomplishments/Strengths	1.737	0.842	.873
Discussion of Teachers' Practice	2.201	0.748	.872
Awareness of Teacher's Practice	2.263	0.991	.909
Summative Assessment of Practice	2.673	0.924	.663
Seeking Teacher's Input	2.730	1.053	.915
Provides for Follow-up	2.903	1.067	.903
Identifies and Addresses Evaluator's Concerns	3.591	1.285	.955

The majority of items from the four sections of the survey significantly loaded onto factors. However, the five outlier items (see previous raw data analysis) from these sections which address APPR requirements did not significantly factor with other items in the sections on approaches, topics, situations and relationships.

One other section of the survey asked participants to report on teacher-to-teacher situations and relationships related to their teaching context that would affect conducting the conference. The items in this section of the survey did load as two factors and did show some reliability as variables, but neither factor correlates with reports of conference effectiveness. This set of non-factoring/uncorrelated items seems to indicate that teachers position what happens between them outside the realm of how evaluators conduct conferences. Since teachers are the focus population who participated in the survey, this outcome for these items is not a surprise when held up against the other factors. This analysis implies teachers do not see what happens between them as part of their performance review. Because this section of items did not contribute significant data to determine factors and/or variables or explain conference effectiveness, I decided to not include this section of items in the over-all data interpretation.

The last section of items focused on asking participants for subjective perceptions about what makes evaluation conferences effective for them. Taken as one group of items when factoring, almost all the items factored with one of three dependent factor variables reflecting measures of effectiveness:

Table 8: Summary of Dependent Factor Variables

Factors: Dependent Variables (measures of effectiveness)	Variable Mean	Variable Standard Deviation	Cronbach's Alpha (Reliability)
Effectiveness=Accountability	2.218	.985	.903
Effectiveness=Cultivating Teacher- Evaluator Relationships	2.356	.996	.934
Effectiveness=Strengthening Teaching	2.731	1.183	.964

As a matter of note, one item from this section of the survey did not significantly load with these factors: “conference effectiveness comes from ensuring teachers conform to district/school policies.” As an outlier, the non-factoring of this item seems to imply participants feel there is little if any connection between compliance to APPR policies and how the conference is effective for strengthening teaching, building relationships, and holding them accountable. The following table provides the correlations between dependent variables that represent effectiveness measures:

Table 9

Correlations of Effectiveness Measures (r) with each other

Measures	1	2	3
1. Strengthening Teaching	1.000	0.851**	0.783**
2. Cultivating Teacher-Evaluator Relationships	0.851**	1.000	0.669**
3. Accountability	0.783**	0.669**	1.000

**significant @ .05 level

The three dependent variables (effectiveness measures) show significant correlation to each other, which implies that the evaluation conference addresses the two purposes of the evaluation process: formative-development and summative-accountability. These correlations are based on how the participants report the evaluation conference contributes to strengthening teaching,

building/cultivating relationships between teachers and evaluators, and holding teachers accountable for student learning and teaching practice.

Once the independent and dependent variables were established, I correlated independent variable factors with dependent variable factors; the dependent variable factors also were correlated with each other. The following table highlights four independent variable factors that correlate the strongest with the effectiveness measures; the four factors and items are listed in order of highest to lowest correlation to the “Strengthening Teaching” effectiveness measure:

Table 10: Factor & Item Correlations with Effectiveness Measures

Independent Factor Variable Correlations with Effectiveness Measures (r)

Factor Variables	1	2	3
Awareness of Teacher’s Practice	.851**	.762**	.675**
Provides for Follow-up	.833**	.703**	.775**
Evaluator’s Fairness	.807**	.781**	.490**
Seeks Teacher’s Input	.631**	.571**	.649**
Discussion of Teacher’s Practice	.613**	.607**	.501**
Attention to Teacher’s Input	.480**	.586**	
Identifies and Addresses Evaluator’s Concerns	.438 *	.232	.484**
Addresses Teacher’s Accomplishments/Strengths	.412*	.503 *	.345*
Summative assessments	.278	.146	.261

Effectiveness measures (r) = **significant @ .05 level; *significant @ .10 level;
Blanks= > .10 level

The four independent factors, evaluator’s fairness, awareness of teacher’s practice, seeks teacher’s input and follow-up/support for improving practice, show the strongest correlations to the formative effectiveness measures (‘strengthening teaching’ and ‘cultivating relationships’) than with the third effectiveness measure (‘accountability’); the third effectiveness measure correlates the most with the factor for providing follow-up and support. The imbalance of the

correlations across the three effectiveness measures for the three noted independent variable factors is also reflected in the individual factor-item correlations with the effectiveness measures. The strength of the correlations implies that participants consider the conferences are effective for strengthening their teaching when the approaches an evaluator uses give each evaluation attention, show an awareness of what teachers are doing, and encourage teachers to collaborate to improve practice. The next strongest factor correlations come under the effectiveness measure for building relationships for the same set of items, indicating that the participants report the effectiveness and/or usefulness of the conference are related to an evaluator's approaches to building interpersonal-relationships. The effectiveness measure of holding teachers accountable shows some significant correlations with the way evaluators show awareness of what teachers have done to improve practice, encourage risk-taking, and invite teachers to develop their own improvement plan.

Four of the other independent factor variables (including all items of the factor) show moderate to weak correlations across all three effectiveness measures despite the high reliability scores. In this group of factors, the strongest correlation is between how evaluators seek teachers' input for addressing concerns and issues about practice. This correlation implies when evaluators use approaches that encourage teachers to collaborate with them on actions to address concerns, teachers consider the conference effective to strengthen their teaching and hold them accountable. What is not indicated by this is why those particular behaviors contribute to effectiveness over the other factors-items that address teachers' accomplishments/strengths, identify their concerns about the teachers' practice, and discuss anything related to compliance, rubric elements and/or reasoning for ratings given to the teacher.

One factor, “Discussion of Teacher’s Practice,” is also correlated with effectiveness measures, but the correlation is problematic. As an independent factor with a high reliability score, the over-all correlations show a minimal-to-weak connection to what makes the conference effective. This factor is problematic because the topic-items, when factored together, are reliable as an independent factor, but then split into sets of topics when correlated with the effectiveness measures. The strongest correlated items coalesce around discussion of student learning and expectations that imply evaluators are focusing on a more formative purpose when including those topics in the conference discussion; whereas the topics that have more summative-rubric focus within the same factor cluster show less-to-no significant correlation to any of the effectiveness measures. The weak-to-no correlations may imply those topics are less important for the evaluators to include in the discussion and do not contribute to making the conference effective for strengthening teaching and building relationships.

The correlation of evaluators’ behaviors (the independent factor variables) with the conference effectiveness measures (the dependent factor variables) show that there are connections between many evaluator behaviors and perceptions of conference effectiveness. All twelve variables were tested to determine which factors contribute the most weight to this perception, and factor variables were entered into SPSS for stepwise regression in order of highest correlation with effectiveness measures. The nine independent variables were tested as groups of “conference behaviors” and “general behaviors,” with the distinction that ‘conference behaviors’ are what evaluators specifically do *during the evaluation conference* and ‘general behaviors’ are what evaluators generally do as a *reflection of their leadership style* when conducting evaluation conferences. Multiple combinations of the two categories of ‘conference’ and ‘general’ behaviors using the four independent factor variables of “fairness,” “awareness,”

“follow-up” and “broad discussion of teaching” showed significant weight over the other five independent factor variables. The following table shows three iterations of testing different variations of the four independent factors:

Table 11: Composite Models Predicting Evaluation Conference Effectiveness

Multiple Regression Analyses Predicating Evaluation Conference Effectiveness

Behavior Variables	Evaluation Conference Effectiveness Measures					
	Strengthening Partnerships		Strengthening Teaching		Promoting Accountability	
	ΔR^2	β	ΔR^2	β	ΔR^2	β
Conference	.595		.603		.390	
Follow-up		.594**		.606**		.333*
Broad discussion		.328**		.328**		.211
Attention to input		.257		.192		
Seeks input						.208
General	.622		.728		.416	
Fairness		.580**		.387**		
Awareness		.268*		.549**		.703**
Conference + General	.651		.770		.490	
Follow-up		.394**		.332**		.247
Fairness		.276		.200		
Broad discussions		.201		.164		
Awareness		.187		.451**		.519**

When testing just factors of ‘conference’ and ‘general’ behaviors, the adjusted R^2 did not show as much weight as when those behaviors were added together. Across the three effectiveness measures, the combination of conference behaviors with general behaviors showed the most accounting for variation for the four factors as a model that predicts conference effectiveness, and that perceptions of effectiveness may be affected by evaluator’s general leadership style combined with what they do in the conference itself. This regression model indicates the over-all effectiveness of the evaluation conference coalesces around evaluators who show encouragement

through follow-up, exhibit fairness of appraisals, have discussions about practice that focus on student engagement, and exhibit an awareness of the teaching context and culture that impacts practice. The strongest predictors, focusing on these factors, possibly contribute to how the evaluation feedback conferences can strengthen evaluator-teacher partnerships and support teacher improvement, reflecting the formative purpose of evaluations over-all. The actions of evaluators that focus on summative purposes such as promoting accountability do not contribute as much to making the conference effective as the other behaviors.

In addition to the scaled items, the survey included two open-ended constructed response prompts that provided participants with the opportunity to elaborate on their individual experiences and/or other information about evaluation feedback conferences. Out of the 39 completed surveys, only a handful of participants elected to submit constructed responses. While this information supports interpretation of the variables and factors, the constructed responses did not contribute to determining the factor/variables. In general, many of the comments, to both questions, help to shed light on how teachers feel about the effectiveness or usefulness of the conference discussion and the approaches used by evaluators when conducting the conference.

The first set of excerpts from the constructed responses focus on issues addressed by the survey items:

- My evaluator recalled the lesson in a detailed manner, showing me that she paid attention to detail during my observation... Very fair and pleasant in her delivery of praises and recommendations.
- The evaluation experience is wide ranging depending upon the evaluator. Seems there is little consistency other than it feels more about bureaucracy than professional development.
- My evaluation conferences largely focus on my planned instructional sequence and pace of the lesson. The feedback I have been given never includes suggestions to improve my instruction. I believe I have been judged fairly. Their feedback generally focuses on my methods to engage and support my struggling students without providing me way to improve my instruction.

Participants' comments indicate the wide range of what happens in the conference with notable mention of the inconsistencies in approaches and over-all experiences that align with the raw data that imply this perception.

The next set of excerpts focus on different aspects of the conferences which items in the survey did not address:

- I have not had very many suggestions made to what I should do differently, so I assume they're satisfied with what they're seeing. If they were not, I might receive more feedback.
- Administrators are overburdened and don't have the time in the day to complete their own work, let alone provide effective feedback on lessons. I find the quality of the feedback to be lacking.
- While I do appreciate the time, effort and energy the evaluator puts into each evaluation- I feel that an increased knowledge of curriculum and skills would allow for more constructive criticism. I will say that I appreciate that my evaluator has done two very thorough evaluations this year, there are some that will not observe their teachers at all. I would rather have the feedback.
- I feel there is a disconnect with the evaluation system partially due to the overwhelming size and condition of the current student population. This forces the duties of evaluation and reflection on teaching to be focused on student behavior more than academic concerns. It creates little opportunities for evaluation discussions that are ongoing and realistic because time constraints are imposed on evaluations.... Evaluations are done more for compliance than a true assistance for teachers.

These comments focus directly on the feedback quality or lack of feedback in general. Since the survey items are designed for objective reporting on behaviors, these comments show the participants' subjectivity related to their individual conference experience which the items did not address.

The comments also show the range for how participants viewed their experiences in general. There are fewer, and more brief, positive comments about the conference and evaluation experience over-all than I expected:

- These conferences are very respectful and informative.
- I feel respected and valued
- Interactions are positive and reassuring
- My post evaluation conference was very positive
- The second observation the evaluator enjoyed the lesson and actually tweeted out a photo...so that was a bit unusual... Interesting feedback of sorts!

Other comments that show a more conflicted, bordering on negative, experience tended to be

lengthier:

- In general, I feel like I have to “get a good score” and that gets in the way of having an authentic conversation about my teaching. I am reluctant to bring up weaknesses and would feel defensive if they were brought up. I look at my evaluator as someone I have to impress and not as a coach who is on my side and wanting me to develop and grow. But I would like to grow and become a better teacher. I trust my evaluator is an excellent educator and has a lot of wisdom to offer me, but don’t feel like the evaluation conference lends itself to that type of relationship.
- My most recent formal observation post-conference was cancelled and never rescheduled despite attempting to meet with her on two separate dates.
- Specific responses to stated lesson objectives are not always discussed. More focus on specific strategies to improve student outcomes and behavioral issues would be welcome.
- When I asked for rationale for certain areas as to why I was not highly effective, the only response I was given was to read the rubric. I have very little interaction with the administrator who observed me.
- This was typical for an evaluation. I don't think they have the proper time, or content-area training to really provide effective feedback.
- The circumstances that surrounded my latest evaluation were impacted by the problems existing outside my classroom that resulted in evaluations done in too short a time frame to be truly an improvement experience for me as a teacher.
- My current lead evaluator is very black and white- there is no room for the human element that is very much a large part of a classroom community. For my evaluator there is never a time when you allow deviation from school rules. We just agree to disagree.

Over-all, the excerpts of the responses reflect a general feeling that the actions of the evaluators impact how effective the conference can be for teachers. As one of the participants wrote, “there is no room for the human element that is very much a large part of the a classroom community,” implying that the focus of evaluators for cultivating relationships and providing formative feedback to teachers suffers at the expense of conducting conferences merely as a way to fulfill the APPR mandates.

Taken as a whole collection, the survey data address the research questions. The independent variable factors address the first research question, “*What are teachers’ experiences with how evaluators conduct evaluation feedback conferences?*” The survey items that address this question reflect the variations in teachers’ experiences but also reinforce what the research literature puts forth as effective protocols for conducting evaluation conferences. The factors related to this question indicate that teachers are aware of those protocols, and report that many of them are happening, but the objective nature of the items did not allow for how teachers *felt* about evaluators taking those approaches and/or focusing on those topics. Even with the variations noted by participants, the findings seem to imply there is a general positivity toward having and wanting evaluation conferences that give useful feedback on practice.

The second research question, “*Is there a connection between the way evaluators conduct feedback conferences, as reported by teachers, and how those conferences affect teachers’ practice?*” can be addressed by the dependent variable factors that provide measures of effectiveness. The rating of ‘effectiveness’ for how evaluators’ approach conducting the conferences for specific outcomes allowed participants to explain for what matters most to them; however, the objective nature of the survey items cannot provide the empirical ‘space’ (other than in the constructed response prompts) for reasons why the participants would feel this way.

The purpose for the third research question, “*Based on teachers’ reports, under what circumstances (if any) is it possible for evaluation feedback conferences to serve both summative and formative purposes?*” was to forefront teachers’ perceptions of the tension caused when evaluation performance appraisals address both purposes. The factors do not indicate a definitive formative-summative line with approaches used or the topics discussed. The only indication of how teachers feel about the formative-summative tension comes through with low or negative factor loadings related to items/variables on rubric and rating discussions, policy compliance, and student test scores (and impact of those scores) on ratings. Other notable variables with little-to-no significance are school context issues associated with teacher-to-teacher relationships. These items did not factor or correlate with the other items, indicating that teachers keep their collegial relationships separate from APPR conferences, even when the evaluator uses an approach that would encourage teachers to seek each other out to collaborate on their practice.

The two sets of data from the survey, one set of statistically determined factors with correlations and one set of qualitative constructed responses, provide findings that over-lap but are not identical. Specifically, the correlations between five factors and the effectiveness measures imply there is a complex understanding of how the conduct of evaluation conferences renders the experience effective for teachers; however, the correlations do not indicate how teachers feel about the particular manner evaluators use the approaches. Correlations can only provide indications, not definitive proof, of which evaluator behaviors and approaches would contribute to conference effectiveness. The constructed responses, even though there are fewer responses than total number of complete surveys, provide only an indication that there are mixed feelings about the conference itself, and there are circumstances that complicate the entire process, not just the evaluation feedback conference. Even though the data sets imply

connections between approaches and effectiveness, the nature of survey research limits how far to interpret those inferences.

The survey data, when combined with RAP interview data, address many of the questions initially brought up on the topic of evaluation conference effectiveness. Because of the limited amount of data, a complete picture of what teachers are feeling still needed corroborating details on how teachers are thinking and feeling about this process. The next research approach, Q Methodology, drew on both data sets and provided a unique way for teachers to voice what they think and feel about evaluation feedback conferences, as well as providing a way to quantify the subjectivity of the teachers that would give the ‘side’ of the issue that is absent in the research literature.

Focus Group: Q-Sort data

After the analysis of the survey data, two small Focus Groups were convened as a method for corroborating or disputing the survey and interview data. The groups were designed to capture teachers’ *subjective* perspectives on how evaluators conduct evaluation feedback conferences. Participants in the groups were given a set of statements, based on the Q-sample of composite statements derived from interview findings and survey items included in common factors, that represent subjective positions/perspectives on the specific approaches and/or topics reported (the *concourse*) to be common across evaluation conference experiences. Based on how participants sorted the q-sample, the concept of *how* evaluators’ approaches make the evaluation conference effective/useful for teachers becomes more defined.

Two small focus groups, one group of nine (including me) tenured/experienced teachers and a second group of four untenured teachers, completed the sorting activity. Each participant was instructed to read the statements with the context that the statements represent what each of

them would say about what makes evaluation feedback conferences effective or useful (i.e. answering the sort prompt). After contextualizing the prompt and explaining how to put the statements into the distribution grid, the groups were given as much time as needed to read and ask clarifying questions about the statements, place the statements into the sort-grid and write individualized comments for items sorted, especially for the statements assigned to the “most” and “least” cells. The following Q-sort distribution grid condenses the raw data (constructed responses were added to the data after factor analysis) from 13 individual participant’s q-sorts (numbers in cells are the q-sample statement numbers) for where each participant assigned the 24 statements; italicized numbers indicate the statements assigned to the cell by untenured participants. The assigned statements in the cells are listed in no particular ranking or order of frequency:

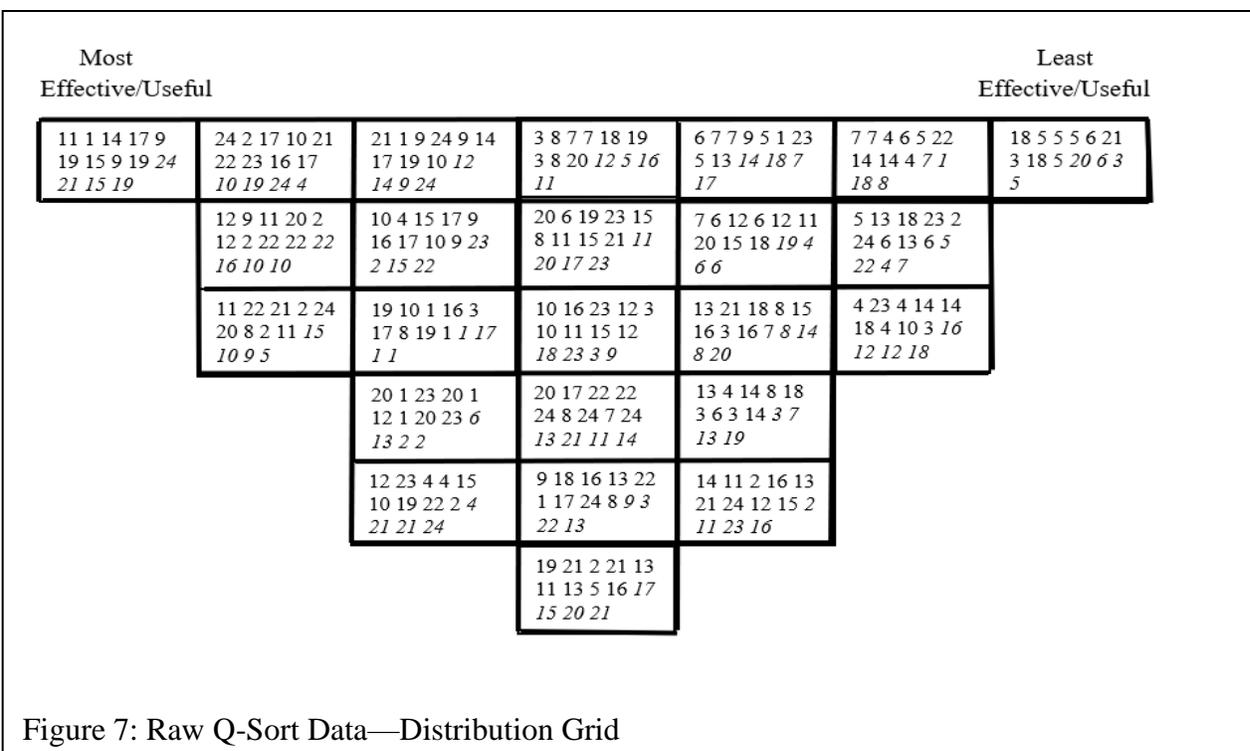


Figure 7: Raw Q-Sort Data—Distribution Grid

The over-all Q-sort grid is notable for how the participants assigned statements in the first/seventh “most effective/useful” and “least effective/useful” cells, showing the range between participants for what statements reflect their strongest perception on evaluators’ approaches to feedback conferences. The second/third and fifth/sixth column of cells represent how participants feel about what is important/unimportant for evaluators to do when conducting conferences and shows how the participants think the approaches have significance, just not the “most” or “least” significance. The middle/fourth column of cells indicate statements that participants consider ‘neutral,’ which is similar to how survey participants could choose “as accurate as not” as a response for some of the survey items. As part of the q-sort process, all statements need to be assigned to a cell, and the neutral column provide space for participants to assign statements in a way that show which statements are least likely to affect their perceptions of effectiveness; neutrally assigned statements does not mean that participants feel the approaches are not significant or important, but view the approaches as what evaluators generally

do as part of implementing the evaluation process that does not have significant impact as compared to the approaches in the “most” and “least” cells/columns.

Before I put the raw data through statistical analysis as per the Q Method approach, I initially coded the collected Q-sorts for the number of times statements were assigned across the cells, especially for “most” and “least,” that would provide general perspectives for what approaches both groups of participants are most sensitive to over-all. The table below highlights what approaches participants felt were “most” effective:

Table 12: Pre-Factored Over-all Most Effective/Useful Statements

Q-Sort Statement	The number of times assigned:			
	Most Effective/ Useful	Effective/ Useful-to Somewhat Eff/Use	Neutral	Less Effective/Useful to Somewhat Less Eff/Use
1. F The evaluator has an idea of what I teach and how I teach in order to discuss his/her concerns, how I will address those concerns, and what I will give the highest priority.	1	5	1	1
9. F The evaluator is open to my opinions, even when they differ from his/her own, which shows he/she is paying attention to what I have to say about my practice.	2	4	1	1
17. S The evaluator discusses my expectations for student learning.	1	5	2	
19. F The evaluator genuinely expresses appreciation for the work I do and acknowledges my accomplishments and strengths that I can build on to improve my practice.	1	4	3	

Note: Statement #; (F)= formative; (S)=summative

As a general feeling and perspective, the two focus groups designate approaches that showed attention, awareness, and acknowledgment of what teachers are doing as significant for making the evaluation conference effective or useful for them. With the exception for when evaluators

discuss student expectations, these approaches are formative for how the evaluators use them to focus on teachers' concerns and opinions, not on the evaluator's summative evaluation concerns.

The next table below highlights what approaches participants felt were the “least” effective:

Table 13: Pre-Factored Over-all Least Effective/Useful Statements

Q-Sort Statement	The number of times assigned:		
	Neutral	Less Effective/ Useful to Somewhat Less Effect/Useful	Least Effective/ Useful
5. S The evaluator discusses all elements of the evaluation rubric the district uses.	1	4	3
6. S The evaluator discusses only the elements of the rubric she/he thinks are important/relevant to my evaluation rating.	1	6	1
18. S The evaluator discusses how students score on assessments of growth, and how I use student information in practice.	2	4	2
7. S The evaluator discusses and explains some/all of the ratings the she/he plans on giving me.	3	5	
13. S The evaluator asks about how I manage my classroom and student issues and interactions with parents/care-givers/guardians (i.e. discipline/positive interactions, contact logs, phone calls, etc) to help make sense of problems/concerns these relationships present in my teaching or other work with students.	3	5	

Note: Statement #; (F)= formative; (S)=summative

The significance of this group of statements is how the majority of the participants (8 out of the 13 for each statement) perceive rubric/rating and student score/behavior management discussions to be “least-to-less-to-neutral” effective/useful approaches for conducting the conference. As opposed to the formative approaches focused on the teacher, these summative approaches are

notable for how the participants feel when an evaluator explicitly uses approaches that focus on observation ratings according to teaching rubrics that provide little-to-no feedback on how to improve practice. Statements #18 and #13, on discussion of student scores and how classroom management impacts practice, respectively, are also notable for how participants feel those approaches are less to somewhat less effective; student scores and classroom management can be considered more summative approaches since the focus is not on instructional practices used by teachers but on external constructs related to the evaluation process that may impact what teachers do in practice.

After coding the raw data, the 13 individual sorts were entered into the software program, PQMethod (Schmolck, 2014). This program combines the q-sorts so that the data can be factored and rotated to produce statistically reliable representations of subjective perspectives of sub groups of participants whose responses are similar. The first step PQMethod takes for statistical analysis is to create a correlation matrix; correlations in Q-Methodology show commonalities across the individual sorts that give some dimension to possible factors that result from factor analysis. As stand-alone data, the correlation matrix does not provide enough statistically significant data to determine factors. However, as the next step in the statistical analysis, PQMethod allows for two options (Centroid or Principal Component) for factor analysis; I purposefully employed Principal Component analysis based on the correlations of all individual sorts for how the factor scores accommodate how the sorts that define the factor based on Eigenvalues that explain the greatest variance but will also be more reliable representation of the shared perspective (Schmolck, 2014). Unlike conventional factor analysis that identifies how individual variables (i.e. survey items or sort statements) cluster together, the focus of statistical analysis in Q-Methodology is to identify which individual participants have similar views that

statistically correlate and then 'load' together on a factor. Q Methodology also highlights the differentiation within and between perspectives by the way the individual q-sorts load together. The q-sorts that 'load' together on a factor represent the defining variables of that factor, and what that factor eventually means comes from looking at how each individual sort in the factor assigns the q-sample statements.

The factor analysis process in PQMethod usually extracts up to eight unrotated factors (the default number in PQMethod program) and provides the Eigenvalues for each factor. Q Research methodology does not propose any explicit number of factors to be used (Eden, et al., 2005; Dziopa & Ahern, 2011; Shemmings, 2006; Wright, 2013), but suggests the number of factors should be based on how the collective sorts that load on factors contribute to understanding the shared perspective represented by the defining sorts. For this study, I determined the number of factors to rotate on Eigenvalues over 1.0 (Eigenvalues at or greater than 1.0 denote strong factors); I also considered the q-sort values of +/- 0.500 to help determine how many sorts have the potential to define a factor. Based on the initial factor matrix of all eight unrotated factors for all 13 q-sorts, I reduced the number of factors to four so that all sorts from the P-set would be included as a defining sort (i.e. factor variable) for a factor.

As the next step in the statistical analysis of the data, I used PQMethod to execute a Varimax rotation in various combinations on the four significantly loading factors. There are two options for rotation in the software: Varimax and by-hand; I employed Varimax rotation since I wanted to avoid 'over-rotating' the factors which has the potential for confounding the factor scores (Schmolck, 2014). Using all four factors for the initial rotation, nine individual sorts loaded between two factors, and two single defining q-sorts loaded respectively on two separate factors; two q-sorts did not load with any of the four factors. A second rotation was performed on

three factors to test if a reduced number of factors would result in all the sorts loading on factors. With this second rotation, seven of the sorts loaded on the first factor, three sorts on the second factor, and one sort on the third; two q-sorts still did not load with any factor. A third rotation with two factors resulted in each individual q-sort loading onto a factor and increased the reliability. In Q-Methodology, the factors can be statistically significant even when there are only two factors since the analysis is primarily concerned with determining the subjective perspectives of small groups of participants. When the factor value was below +/- 0.500, I assigned the sort to the factor for which it had the highest value between the two (particularly sorts #11 and #12). Even with the weak values, the sorts can provide interpretive data to help distinguish the attitudes of the two factor groups. The following table shows how each sort loads as a 'defining' sort (defining sorts are bold/highlighted) for a factor as well as the Eigenvalues, percent of variation explained, standard deviation, standard of error, and composite reliability for each factor:

Table 14: Factor Loadings with Defining Sorts and Factor Reliability

Summary of Q Data Factor Analysis

Factor Loadings based on Defining Q-Sorts			
Q-Sorts	Authentic and Values Teachers (n=9)	Accurate Understanding and Useful Insights (n=4)	
1	.770		-.117
2	.655		.372
3	.545		-.579
4	.070		-.566
5	.660		-.038
6	.563		.323
7	.577		.298

8	.675	.040
9	.896	.0120
10	.785	.082
11	.448	-.012
12	.119	.448
13	.178	.876
Eigenvalues	4.60	1.91
% Var	35	15
<i>SD</i>	1.504	1.504
<i>SE</i>	.164	.243
α	.973	.941

Once the sorts were factored, PQMethod then provides factor arrays that represents how statements from the defining sorts for the factor collectively distribute across the distribution grid, thus creating an ‘ideal’ sort that can be interpreted as the shared viewpoint of the participants that load with this factor. To create the factor array, the program averages the placement value (+3 to -3) for all the statements of the defining sorts, which are labeled “Z-Scores.” Z-scores reflect the shared ranking of the statement by the individual participants flagged as defining sorts for the factor, not identical rankings. These scores rank the statements in descending order of significance for the factor, flagging statements with Z-Scores +/- 1.5 as significant exemplar statements. A Z-score within +/- .01 indicates placing a statement in a ‘neutral’ position in the grid; in Q Methodology, ‘neutral’ does not mean that a participant feels the statement does not reflect having a value, but more along the thinking that if the action the statement represents happens or not, there will be no impact on effectiveness level.

I used a combination of the factor arrays with constructed responses from the flagged sorts to determine how each factor represents a collected perspective that would help define how

teachers feel about the evaluators' approaches to conducting evaluation conferences that are effective and useful.

Table 15

Summary of Q-Sort Statement Factor Scores (Z-Scores) with Corresponding Ranks within Factors

Q-Sort Statements	Authentic and Values Teachers		Accurate Understanding and Useful Insights	
	Z-Score	Rank in Factor	Z-Score	Rank in Factor
1.The evaluator has an idea of what I teach and how I teach in order to discuss his/her concerns, how I will address those concerns, and what I will give the highest priority.	.78	8	.49	8
2.The evaluator lets me identify aspects of my teaching that I consider to be areas of concern, how to address those concerns, and what priority I should address those concerns.	.79	7	-.93	21
3.The evaluator lets me know before the conference anything he/she wants to discuss and explain what he/she wants to accomplish in our conference.	-.97	20	-.76	19
4.The evaluator asks me about what to look for in the observation.	-1.44	23	.44	9
5.The evaluator discusses all elements of the evaluation rubric the district uses.	-2.27	24	-.77	20
6.The evaluator discusses only the elements of the rubric she/he thinks are important/relevant to my evaluation rating.	-1.33	22	.81	6
7.The evaluator discusses and explains some/all of the ratings the she/he plans on giving me.	-.95	19	-1.21	22
8.The evaluator encourages me to suggest options for addressing both of our concerns as well as collaborate on coming to consensus on how to address concerns both of us can support.	-.01	14	-.68	16
9.The evaluator is open to my opinions, even when they differ from his/her own, which shows he/she is paying attention to what I have to say about my practice.	1.13	2	-.48	15
10.The evaluator takes the time to give my evaluation individual attention so that the discussion focuses on useful feedback.	1.04	3	1.44	2

11.The evaluator discusses whether/how my students are actively engaged in learning/ meeting learning objectives and how I assess their learning.	.94	5	.12	12
12.The evaluator focuses the conference on how I plan lessons/adapt instruction for different students.	-.29	16	-.28	13
13.The evaluator asks about how I manage my classroom and student issues and interactions with parents/care-givers/guardians (i.e. discipline/positive interactions, contact logs, phone calls, etc) to help make sense of problems/concerns these relationships present in my teaching or other work with students.	-.72	17	.36	11
14.The evaluator encourages me to develop relationships with other teachers who share the same concerns as I do or seek out other teachers for advice on how to address some of the concerns/issues brought up in the conference.	-.77	18	-.75	18
15.The evaluator focuses on/discusses in-depth useful suggestions for specific actions I might take to change my teaching, including new ideas that may mean making mistakes, with what support he/she will give me based on what we decide I will try to do.	.13	12	1.41	3
16.The evaluator asks me to come to the conference prepared to discuss anything I think needs attention, including my goals, concerns and/or something from the observation.	-.06	15	-1.49	23
17.The evaluator discusses my expectations for student learning.	.99	4	-.32	14
18.The evaluator discusses how students score on assessments of growth, and how I use student information in practice.	-1.00	21	.44	10
19.The evaluator genuinely expresses appreciation for the work I do and acknowledges my accomplishments and strengths that I can build on to improve my practice.	1.74	1	-.73	17
20.The evaluator is respectful sharing insights/opinions of teaching and asks for advice on how to address issues that affect me.	.36	10	-1.78	24
21.The evaluator uses accurate information when discussing what was observed and shows an awareness of what I have done to improve teaching.	.36	11	2.06	1
22.The evaluator shows that he/she is a good judge of my effectiveness because he/she regularly observes my teaching, understands the curriculum I follow and understands instructional challenges I face.	.94	6	1.12	4
23.The evaluator conveys a clear vision of what she/he wants students and the school to accomplish, showing that she/he has the interests of students in mind.	.05	13	.64	7

24. The evaluator focuses on using the conference to help me see situations related to my teaching from different vantage points, offers useful perspectives on things she/he observed in my teaching, and identifies ways to further my professional development on those situations and my overall practice.

.57 9 .85 5

Note: Z-Scores reflect shared ranking of statement; highlighted Z-SCR +/- ≥ 1.5 flagged as significant.

The first factor flagged nine out of the thirteen q-sorts as ‘defining’ sorts. The factor array indicates this group of participants put a priority on wanting to feel valued and respected for being professional and wanting the evaluator to show he/she knows what is happening in the teachers’ practice

Table 16: Q-Sort Factor Array 1

Most Effective/Useful							Least Effective/Useful	
+3	+2	+1	0	-1	-2	-3		
19	9	1	8	3	4	5		
	10	2	15	7	6			
	17	11	16	12	18			
		22	20	13				
		24	21	14				
			23					

Q-Sort Factor Array 1: Summary of Top-Ranked Defining Statements (n=9)

Statements	Z- SCR	Q-SV
19. The evaluator genuinely expresses appreciation for the work I do and acknowledges my accomplishments and strengths that I can build on to improve my practice.	1.74	3
9. The evaluator is open to my opinions, even when they differ from his/her own, which shows he/she is paying attention to what I have to say about my practice.	1.13	2
10. The evaluator takes the time to give my evaluation individual attention so that the discussion focuses on useful feedback.	1.04	2
17. The evaluator discusses my expectations for student learning.	.99	2
23. The evaluator conveys a clear vision of what she/he wants students and the school to accomplish, showing that she/he has the interests of students in mind.	.05	0
8. The evaluator encourages me to suggest options for addressing both of our concerns as well as collaborate on coming to consensus on how to address concerns both of us can support.	-.01	0
7. The evaluator discusses and explains some or all of the ratings the she/he plans on giving me.	-.95	-1
3. The evaluator lets me know before the conference anything he/she wants to discuss and explain what he/she wants to accomplish in our conference.	-.97	-1
18. The evaluator discusses how students score on assessments of growth and how I use student information in practice.	-1.00	-2
6. The evaluator discusses only the elements of the rubric she/he thinks are important or relevant to my evaluation rating.	-1.33	-2
4. The evaluator asks me about what to look for in the observation.	-1.44	-2
5. The evaluator discusses all the elements of the evaluation rubric the district uses.	-2.27	-3

Note: Q-SV= score values of +/- 3 including 0; Q-SV=0 indicates 'neutral.'

Constructed Responses for “most” effective/useful statements:Written comments on cards:

Statement 9: “values me”

Statement 10: “time to ask questions”

Statement 17: “love this”

Statement 19: “Expressing appreciation shows that she noticed what I did and that I’m doing something right. Showing I need strengthening is what should be happening. I want to know what I can do next.”

Specific sort comments on statements 9, 10, 19:

Sort #5: (17) “Using expectations for student learning that are articulated by the teacher sets a more meaningful standard for evaluation, for both sides of the equation.”

Sort #7: “These all address a collaborative way of improving based on my needs, my students’ needs, my school’s needs, NYS needs, and the evaluator’s needs to be successful. These are effective!”

Sort #8: (19) Needs to hear the positive—does like to hear ‘This is good and this is what to do next’/items show “they treat me as a professional but also as a learner. I can grow and won’t be offended by suggestions. This is how I’d evaluate someone.” All the items are “respectful”

Sort U#1: (19) They appreciate the effort and understand (9, 10) Support and understanding

Sort U#2: Helps me see everything in different ways (19, 10) Evaluator shows he/she cares (9) That’s what matters most, but not most effective

Constructed Responses for “least” effective/useful statements:

Written comments on cards:

Statement 4: --“I know what’s important to my lesson” “don’t bother coming in if you don’t know what to look for.” “I feel like that would give me an advantage. As teachers we usually know what they (evaluators) are looking for so if I work on those key elements, and let them know to look at those things, then in reality I would do better.”

Statement 5: “Doesn’t help me at all with teaching practices. I can read it myself. I don’t want to waste my time with this when I can discuss actual teaching.”

Statement 6: “doesn’t even know the whole rubric” / “leaves out important stuff”

Statement 18: “not a full pic of a student’s learning”

Specific sort comments on statements 4, 5, 6, 18:

Sort #1: (4, 5) “Rubric questions—not a big deal to me”

Sort #5: (4) “... basically ask teachers to evaluate themselves. This is something good teachers do anyway, but it is difficult to share your own insights with an evaluator who may use that information against you.” (6) “The evaluation process has tremendous ____ for subjective bias, and this statement is a perfect example of how that works.”

Sort #6: (5) “The rubric does not measure/evaluate affective teaching. Which leads to effective teaching, in my opinion. Someone who is not in the profession developed the rubric—not meaningful for me. These issues do not come into the conversation during the evaluation conference.”

Sort #7: “These focus on rubrics, ratings, and preparing for observations in order to get the best score I can. These do not focus on how effective my teaching practices are or how my students are learning. These are ineffective!”

Sort #8: Taken as a whole ‘less-to-least’ effective/useful items represent “a waste of time”—“I would be annoyed to get called out of my classroom to listen to this” (5) “don’t want to waste my time” ‘I can read it’ “waste of time giving me info I should already know”

Sort U#1: (5) I don’t like the reliance on rubric. Check box (4) Cookie cutter (6) Check boxes

Sort U#2: (4) These are minor things, but not less effective (6) Everything is important

Over-arching comments:

Sort #1: statements on positive side show “most important are the genuine discussions about my practices/strategies” “Admin needs to be in-tune with their people and in touch with what they are teaching.” “Do they know me/ their people?”

Sort #5: the ‘most’ side statements “support a professional, objective and constructive conversation that focuses on meaningful dialogue regarding measurable parameters.”

The top ranked items emphasize that the evaluator has a sense of respectful and authentic appreciation for the teachers’ efforts, based on an understanding of what the teacher is doing. Teachers value this give-and-take represented in the comment, “values my opinion but also has an opinion” (Sort #8). The regular observation and focus on students contribute to the perception that the evaluator is showing he/she values and respects the teacher, that there is an openness to the feedback given in such a context. When the evaluator is more authentic and the focus is on students, then feedback becomes a “more collaborative way of improving” based on the needs of the teacher, students, school, NYS, and “evaluators’ need to be successful” (Sort #7).

Just as authenticity and making the teacher feel valued promotes feedback efficacy, discussing all the elements of the evaluation rubric are expressly and explicitly noted for being “a waste of time,” especially when the evaluator discusses only the elements he/she thinks are important/relevant to the evaluation rating. Teachers shared the perspective that such a limited summative approach is “not a positive approach.” The feeling of being valued and appreciated that shows authenticity is undermined and becomes the antithesis of efficacy by focusing on rubrics and ratings.

The individual q-sorts (representing individual teachers) which share this perspective are not related to the number of years teaching. Across the years of experience, the highest loading sort is from a 22-year experienced teacher, and the next highest is from a third year, untenured teacher. The other untenured teacher (with this year being the first year of teaching) was the lowest loading on this factor but still made significant remarks that resonate with the other lower loading sorts from 31-year and 16-year experienced teachers. Over-all, this factor focuses on how the evaluator makes the teachers feel and the time the evaluator takes to make them feel this

way. This group's perspective reflects how important it is for the evaluator to be authentic and show the value for how the teachers feel about their own practice.

The second factor flagged four out of the thirteen q-sorts as 'defining' sorts. This group of participants put a priority on evaluators who use approaches that show an accurate understanding of their practice that is reflected in the feedback that is useful:

Table 17: Q-Sort Factor Array 2

Most Effective/Useful +3		+2	+1	0	-1	-2	Least Effective/Useful -3	
21	10	1	9	3	2	20		
	15	4	11	5	7			
	22	6	12	8	16			
		23	13	14				
		24	17	19				
			18					

Q-Sort Factor Array 2: Summary of Top-Ranked Defining Statements (n=4)

Statements	Z- SCR	Q-SV
21. The evaluator uses accurate information when discussing what was observed and shows an awareness of what I have done to improve teaching.	2.062	3
10. The evaluator takes the time to give my evaluation individual attention so that the discussion focuses on useful feedback.	1.443	2
15. The evaluator focuses on/discusses in-depth useful suggestions for specific actions I might take to change my teaching, including new ideas that may mean making mistakes, with what support he/she will give me based on what we decide I will try to do.	1.411	2

22. The evaluator shows that he/she is a good judge of my effectiveness because he/she regularly observes my teaching, understands the curriculum I follow and understands instructional challenges I face.	1.121	2
24. The evaluator focuses on using the conference to help me see situations related to my teaching from different vantage points, offers useful perspectives on things she/he observed in my teaching, and identifies ways to further my professional development on those situations and my over-all practice.	.849	1
6. The evaluator discusses only the elements of the rubric she/he thinks are important or relevant to my evaluation rating.	.808	1
11. The evaluator discusses whether or how my students are actively engaged in learning and/or meeting learning objectives, and how I assess their learning.	.117	0
5. The evaluator discusses all the elements of the evaluation rubric the district uses.	-.773	-1
2. The evaluator lets me identify aspects of my teaching that I consider to be areas of concern, how to address those concerns and what priority I should address those concerns.	-.925	-2
7. The evaluator discusses and explains some or all of the ratings the she/he plans on giving me.	- 1.210	-2
16. The evaluator asks me to come to the conference prepared to discuss anything I think needs attention, including my goals, concerns and/or something from the observation.	- 1.487	-2
20. The evaluator is respectful sharing insights/opinions of teaching and asks for my advice on how to address issues that affect me.	- 1.785	-3

Note: Q-SV= score values of +/- 3 including 0; Q-SV=0 indicates 'neutral.'

Constructed Responses for "most" effective/useful statements:

Sort #2: (22) "If they paid more attention to these items, I believe the evaluation process would be much more effective-especially for younger teachers."

Sort U#3: (15) The evaluator is telling me specifically what they are looking for, for my next assessment (22) same person is reliable

Sort U#4: (21) Fundamentally, it seems an observation is most effective when it yields feedback that is: a) accurate, so that feedback is relevant, b) specific, so that I can tell how to improve with specific steps, c) substantive, based on the evaluator's experiences, goals for the building

Constructed Responses for "least" effective/useful statements:

Sort #3: (7) "Tell me about my teaching. Don't justify your ranking. But if you explain how I could improve score and teaching then okay."

Sort U#3: (7) 'calling too much attention to what's wrong'

Sort U#4: (7, 16) "they don't need to ask me" (20) Results that improve my practice are more important than my personal feelings/reaction in that moment. If I am able to remain receptive to the feedback, I would rather have an honest, accurate, disrespectful evaluator than a respectful, but less effective one.

The defining q-sorts for this factor are notable for the opposing factor loadings; two sorts from similarly experienced teachers are negatively set with two sorts from untenured second year teachers. Even though the sets of teachers seem to be in opposition, all four sorts indicate that feedback conferences are effective/useful when there is a sense that evaluators have regularly been in teachers' classrooms, not just for the formal observations. When teachers perceive evaluators are doing this, the conference becomes effective and useful because evaluators show an awareness of what is happening over-all in the school and have a sense of what is important in the curriculum, classrooms and school in general. When teachers feel that evaluators are investing this time to get a sense of what is happening in teachers' day-to-day practice, then what the evaluators say and do in the conference will reflect an accurate understanding of their practice. The teachers who share this perspective place more emphasis on the authenticity of what evaluators have to say about their practice based on accurate *understanding* of what they do rather than evaluators giving *appreciation* for what they do.

The confidence teachers feel about the evaluators' understanding contributes to how open teachers are to the evaluators' feedback for the purpose of improving practice. The teachers who share this perspective feel there is a lack of substance behind the evidence evaluators use based on inaccurate/inauthentic evidence when evaluators come in for only a one-time-fits-all required observation, which impacts how teachers perceive the feedback given during the conference. When evaluators show a sense of understanding the over-all instructional context of what teachers are doing, as well as the steps taken by teachers to improve practice (with or without the evaluators' suggestions), teachers feel there is substance to the feedback that makes the conversation relevant. When the teachers feel the evaluators are using the conference time to justify what is on the rubrics from their position as evaluators, or the evaluators are seemingly

unaware of what the teacher does in the day-to-day context of teaching, then teachers consider the effectiveness or usefulness of the feedback conference to be negated by the lack of awareness. As one of the untenured teachers notes, “I would rather have an honest, accurate, disrespectful evaluator than a respectful, but less effective one.”

Whether the factor array shows evaluators’ approaches are effective/useful when authentic and respectful or when accurate and useful, both factor arrays share a significant defining statement, #10, that focuses on how “*The evaluator takes the time to give my evaluation individual attention so that the discussion focuses on useful feedback.*” For an evaluator to be either authentic or accurate, the element of time seems to be significant for the teachers. In the over-all group discussion, participants noted multiple times that authenticity and accuracy need time to develop as part of the evaluators’ skill-set for making the feedback useful and the conference itself worthy of the time it is given, even if it is brief. The group, as a whole, also noted that this kind of ‘time’ is a luxury in the current climate of getting as many evaluations done in the most expedited time in order to fulfill the accountability requirements for APPR. Even though this element of time is noted as what is effective/useful to teachers for a feedback conference, the group acknowledged it reflects more of a ‘wish’ than a matter of ‘fact’ that it happens.

Just as time seems to be a shared concept that contributes to the over-all effectiveness of the feedback conference, both factor arrays note rubric discussions/elaborations to be least-to-less effective approach to feedback conferences, albeit not the same exact statement(s). The general perspective seems to be when evaluators rely on using a rubric lens when conducting the conference that constricts the scope and depth of the discussions to only summative ratings, rubric dimensions and ‘check-off’ items from the APPR plan in place, teachers lose interest in

what the evaluators have to say because of the ‘cookie-cutter’ feel. The group discussion noted that rubrics and ratings are a necessary part of APPR, and evaluators have to adhere to the way the district wants the rubrics applied/ratings determined. That being said, the group also discussed the ability of individual evaluators to make that part of the conference discussion work for both them and the teachers. When evaluators do not have that ability and rely on discussing ratings/rubrics because it is an easy ‘script’ to follow and fulfill the ‘requirements,’ then teachers do not have an investment in what the evaluators have to say. A general feeling in the group was that an evaluator who can bend the rubric/rating discussion into something authentic/accurate would be approaching the conference in a way that would make it effective/useful for any teacher at any level.

Summary

Three distinctly different research approaches, yet common themes emerge:

- *Evaluators’ Awareness and Understanding* of what teachers are doing in practice, how/why teachers make decisions in practice, and what struggles/challenges/curriculum issues teachers face day-to-day
- *Evaluators’ respect* for teachers’ time, opinions, and reflective/reflexive judgment of their own practice
- *Evaluators’ Fairness* when applying standards of teaching, gathering evidence used to determine evaluation scores/ratings, and understanding what really counts in each individual teacher’s professional context that impacts what is done day-to-day

These themes resonate within the context of each set of data. A surprising element across all three sets of data is how teachers look to, and even expect on some levels, evaluators to be as honestly reflective/reflexive about the purpose/conducting of the conference as the teachers are about their practice. It does not seem to be a matter of simply ‘trusting’ the evaluator to be aware, respectful, fair, and equitable; there is a sense that teachers want and expect evaluators to have a level of self-awareness about their role in the whole process.

Each participant in the interviews, the survey, and focus groups also brought the individual ‘human element’ into the mix. The interview participants brought out stories of ‘good’ evaluation experiences that were coupled with a ‘nostalgic’ element of what once worked. In a similar way, participants in the focus group struggled with the statements as representing what they ‘wish’ would happen in conferences, and in fact, what the statements said often do not happen. Some of the survey constructed responses also noted what was ‘supposed’ to happen may not always be what ‘does’ happen. As one of the survey participants noted in a constructed response, looking at the evaluation conference experience from the perspective of the survey items made him aware of how ‘complex’ this issue of evaluators’ approaches is for teachers and administrators.

A related theme to this perspective that emerged is the concept of *time* which evaluators have and take with each evaluation. Whenever participants from any of the research activities mentioned time evaluators take for observations and conferences, I noticed an accompanying comment about external constructs (i.e. APPR mandates, too many evaluations done by single evaluator, complicated procedures) having an impact on this approach. The focus group participants all put the statement about taking time with individual evaluations on the effective/useful side of the grid, but this statement was not explicitly asked about in the interviews or survey. However, I notice that time for evaluators to do thorough and multiple observations was a recurring theme in the RAP interviews, and one of the constructed comments from the survey mentions an element of time for doing observations has impacted the thoroughness of the evaluation over-all.

Each set of data, collected in methodologically different ways, brings out the nuances and what is unique about individual experiences for how teachers perceive feedback conferences,

revealing the complexity of the phenomenon. As strictly qualitative data, the RAP interview narratives show teachers' frustration with a process they consider flawed when they cite inconsistencies in evaluation approaches and feelings of powerlessness over their professional narratives which are based on flawed processes. These frustrations come out in the interview narratives with an undercurrent of negativity about how evaluators, and districts in general, implement the procedures for evaluations. Because the participants were in one-on-one interviews, they seemed open to give their unvarnished, subjective perceptions of evaluation situations and contexts in their 'own words' as opposed to survey items (and q-sort statements to a certain extent) that are not.

As strictly quantitative data, the responses to all objective sections of the survey show the varied nature of participants' experiences with how evaluation conferences are conducted. The survey gave participants an opportunity to respond to a wider array of evaluator behaviors and note which ones, through indirect correlations, were associated with each other and then with assessments of effectiveness. What emerged are distinctions between conference behaviors and general leadership approaches that are employed, simultaneously, in evaluation conferences. *How* evaluators employ those approaches during the conference shows indirect correlations suggesting that there are connections between the conduct of the evaluation process and what makes the conferences effective for strengthening their teaching, building relationships and holding teachers accountable. As opposed to the direct nature of asking interview and focus group participants to describe experiences and make connections between approaches and effectiveness, the survey only allows for inferring such connections. The survey did provide the opportunity for making connections between approaches and effectiveness in their 'own words' with the open-ended constructed responses, and the majority of the participants who did respond

to the open prompts were more positive over-all than interview participants. However, there is no way to directly ask/know what contributes to this underlying positivity.

The mixed nature of the Q-approach allows subjective/objective data from the focus groups to open the 'black box' of what teachers are thinking and feeling about evaluation conference experiences and how evaluators approach those conferences. The two perspectives that emerged from the Q approach align with both interview and survey data, but also reveal subjective perspectives that survey data does not supply, and statistically validate those perspectives which cannot be done with narrative data from interviews. The two perspectives revealed by the Q research data highlight how some teachers are more sensitive to some considerations than others, namely that effective conferences are when evaluators approach conferences authentically and respectfully or show accuracy and provide useful feedback and suggestions. Both perspectives resonate with interview data and survey responses, but on a more specific level because the Q-approach centers on directly asking teachers to focus on those feelings. This specificity about what teachers think and feel provides definition and clarity to the over-all issue concerning what approaches make evaluation conferences effective for addressing dual functions and purposes.

Chapter 5

Discussion—Conclusions

Teacher performance evaluations have become an experience that elicits a broad range of emotions and attitudes from both teachers and administrators-evaluators about how evaluations are implemented according to NYS Education Law §3012-d (APPR law). What most educators seem to be feeling is a tension with how the law, and therefore most/all districts' evaluation systems, imposes regulations and mandates on how to use teachers' evaluation ratings for dual functions and purposes. This tension comes from how teachers and administrators/evaluators struggle with balancing functions of supervision and evaluation for formative and summative purposes all under the one, over-arching system; however, school organizations are finding achieving this balance difficult if not impossible (Popham, 2013) when the law seems to guide districts, intentionally or not, almost exclusively toward using summative evaluation ratings for accountability functions and purposes (NYSED, Guidance Document, 2018).

The purpose of this study has been to explore what lies at the center of this tension by collecting data on teachers' experiences with the one component of evaluations where this tension is most evident, the observation feedback conference. Despite the challenges presented by the low numbers of participants at each stage of this study, some interesting factors emerge within and between the data sets that provide directions for further research. The data collected in three methodologically different ways reveal similar yet complex affective perceptions held by teachers for how evaluation feedback conferences are or are not effective for them based on how the evaluators conduct those conferences. However, even though there are similar perceptions of which factors are important, participants across the data sets diverge on how evaluators use those noted approaches to address the functions and purposes of the evaluations during the feedback

conferences, with some participants taking on a more critical tone and others a more positive tone. Understanding the affective responses, from both the critical and positive positions, and how those responses contribute to the various dynamics of evaluation feedback conferences, has the potential to move the field toward developing a theory related to how an evaluator's approach to conducting evaluation feedback conferences contributes to the effectiveness of the evaluation process.

Research Findings

An integral characteristic of mixed methods studies is how the research blends the data collection methods for addressing the research questions and provides insights that reflect a fuller picture of the topic under study. As part of this research study's design, the three sets of data specifically address the first two research questions focused on teachers' lived experiences and sense-making of feedback conferences, particularly focusing on teachers' reports for what approaches and behaviors evaluators use that make the evaluation feedback conferences effective for impacting practice. What the data sets reveal are teachers who experience common evaluator behaviors, across and within districts and/or schools. The predominant evaluator behaviors are notable for how they affect interpersonal relationships between teachers and evaluators, which teachers perceive (positively and negatively) as connected to the evaluators' leadership 'style' that, in turn, impacts the way the feedback conference is conducted and for having an impact on their practice. Teachers' reflections on their experiences also are notable for common external contextual constructs associated with feedback conferences, such as the APPR law and policy/accountability mandates, which complicate how evaluators conduct feedback conferences, and in turn impact the effectiveness of evaluation conferences for the teachers.

As for the third research question, investigating the possibility that feedback conferences could serve dual formative and summative purposes, the collected data are too limited to make confident generalizations or inferences. However, the limited data do reveal some dimensions of conference effectiveness teachers recognize, albeit indirectly, as having an impact for how feedback conferences could/can, indeed, address dual purposes. Further research is needed of teachers' perceptions on those dimensions, specifically, to fully address the issue of the tensions felt by teachers when they perceive feedback conferences are used for one purpose over the other. The following discussion examines each data set for insights on the factors, constructs and dimensions related to all the research questions:

Narrative Data—Semi-Structured Interviews

As the initial stage of this study, the Research Apprentice Project (RAP) examines the impact of evaluation systems on teachers' practice, focusing specifically on the feedback from evaluators. I interviewed a small sample of veteran teachers (between 15 to 25+ years of teaching) who recounted professional experiences with fluctuating evaluation systems and reflected on the impact various iterations of performance evaluations have had on their practice. The transcripts and fieldnotes from these interviews capture participants articulating a sense of understanding themselves as teachers and their practice throughout their professional lives, but their narratives also imply feeling professionally 'voiceless' for expressing that self-awareness within the context of performance evaluation system (i.e. mandated state APPR law) or even in the evaluation conference.

Even though eight narratives are not enough to generalize across an entire population (i.e. teachers in NYS who are evaluated with §3012-d), the themes that emerge from the narratives shed light on issues and tensions that teachers feel when asked about their experiences with and

during evaluation feedback conferences. Discourse analysis of the narrative data shows participants are especially focused on the authenticity of evaluators' appraisals based on the interpersonal relationships established by the evaluators within the context of feedback conferences. The narratives are critical of evaluators who have not established a collaborative and equitable relationship outside of the conference context, which then impacts the relationship within the context of the conference. The evaluators' interpersonal approaches then create tension over the trust-power/control of the individual evaluations as teachers' professional narratives and teachers' perceptions of self-worth reflected in evaluators' assessments of practice.

The interview narratives also express perceptions of how underlying external constructs (i.e. school culture and context, organizational oversight of evaluation processes, and APPR legislation) impact evaluators' approaches to conducting feedback conferences, which then also impacts the effectiveness and utility of the evaluation conference as part of the process over-all. The participants do not necessarily share the same experiences with these constructs, but in some way each participant does articulate that culture, context, policy, and politics affect the effectiveness of the conference experience. Interview participants experience similar contextual issues, such as changes to the APPR law, policies and regulations, that are out of teachers' and evaluators' control, but they note how approaches do vary between individual evaluators when they conduct conferences under those confines and contexts. The participants' narratives of conference experiences attribute specific approaches to be rooted in evaluators' leadership styles which they 'follow' when implementing APPR mandates that adhere to the contextual evaluation constructs inherent in the APPR law. Teachers note how evaluators using approaches that seem to fulfill a 'check list' of APPR requirements (reflecting only a summative approach) over the

necessity of establishing a collaborative and equitable professional relationship (reflecting a formative approach) to address policy or APPR mandate requirements does impact the effectiveness of the conference and even the evaluation process over-all.

Further discourse analysis of the interview data reveals themes related to lack of agency for determining teaching efficacy within individual teaching contexts and the questioning of professional judgment that emerged from performance evaluations. I notice participants seem to feel evaluations do not reflect authentic teaching contexts or how they identify as teachers within those contexts. These participants perceive the evaluation process over-all, not just the conference, reflects a narrative about them as teachers. The critical nature of the participants' perceptions of evaluators' approaches came through the narrative data when they note how they felt evaluators did not *know* them or their teaching, which then impacts the effectiveness or utility of the feedback conference. Their narratives intertwine themes of interpersonal relationships, professional efficacy, and personal reflectivity, as well as reflexivity, which participants note as what makes them identify as 'teachers,' but feel that the evaluation process, most notably the feedback conferences, diminishes and does not account for what they feel they actually do in the classroom day-to-day.

The RAP narrative data from the teachers' perspectives present a more complex perception about what makes evaluation conferences effective than what the literature and research propose. When I compare the themes of self-assessment and marginalization that emerge from the interview narratives to the background research literature on educational personnel evaluation and supervision, I notice the literature that discusses teacher evaluation processes typically ignore *teachers' perspectives* on how evaluation feedback conferences impact teachers' practice. When the discussions do focus on conferences, they focus on

educational leaders' attitudes and what strategies they should use to conduct feedback conferences, not on how teachers subjectively feel about how conferences are conducted. The research literature also focuses on how administrators who are evaluators should approach the conference to elicit the teachers' investment in their own practice, but not how teachers feel those approaches contribute to the utility or effectiveness of the conference.

What I notice from reanalyzing the RAP interview data becomes incorporated into the next step of my dissertation study. The themes from the RAP interviews provide the framework for extending and furthering the research on the effectiveness and usefulness of the feedback conference from the *teachers'* perspectives. Taken as shared perspectives on a common experience to unpack, I mainly focus on teachers' specific feelings towards how evaluators conduct feedback conferences (especially interpersonal feelings towards specific evaluators and trust) and mandated requirements for APPR ratings related to the culture and context of their individual teaching situations. Those subjective narrative themes give voice to teachers' feelings that are absent in the literature for understanding how the feedback conference can address issues with conducting APPR evaluations over-all.

Survey Data

As strictly quantitative data, the responses to the objective sections of the survey show participants share common or similar experiences with how evaluation conferences are conducted across districts, even though there is a low number of responses to the survey. The 39 complete surveys are not enough to make broad generalizations across the sample population of teachers in NYS, but the statistical analysis of the data set does show that the teachers surveyed experience common approaches and complicating contexts that impact how effective conferences are and how they affect their practice. Since the survey collected objective reports of

experiences, any personal perspectives and/or affective responses can only be inferred through correlation with subjective assessments of conference effectiveness and the very limited subjective constructed responses on those experiences. That being noted, what the collective survey data show are teachers reporting experiences with evaluators who are attending to many of the expected topics related to general evaluation processes by using a number of specific approaches suggested by the literature on educational leadership, supervision and evaluation. The over-all, most common factors that emerge from the objective responses imply evaluators should include approaches reflecting they are aware of teachers' practice, fair in the appraisal of practice, and include teachers in the process in order to strengthen their teaching and develop collaborative relationships. These objective (independent) factors are strongly correlated with teachers' subjective reactions to those reported actions and behaviors, which contribute to teachers' perceptions of conference effectiveness.

After completing a statistical analysis on the objective data that included performing PAF and Cronbach's Alpha reliability checks, I conclude there are nine independent variables representing evaluators' behaviors and approaches related to the three dependent factors measuring conference effectiveness. Out of the nine independent factor variables, the most notable objective variables that emerge from the survey data coalesce around how evaluators approach conducting the conference, involve the teachers in the evaluation process/conference, include specific topics as protocol for/during the evaluation process, and develop relationships with teachers that impacts conference effectiveness. These four factors show stronger correlations to the first two effectiveness measures ('strengthening teaching' and 'cultivating relationships') than with the third effectiveness measure ('accountability'). The strongest correlations imply that participants think conferences are effective for strengthening their

teaching when the evaluator gives each evaluation attention, shows an awareness of what teachers are doing, and encourages teachers to collaborate to improve practice. The next strongest factor correlations come under the effectiveness measure for building relationships for the same set of items, indicating that the participants report the effectiveness and/or usefulness of the conference are related to an evaluator's approaches to building interpersonal-relationships. The third effectiveness measure focusing on accountability correlates the most with the independent factor for providing follow-up and support. The effectiveness measure of holding teachers accountable also shows some significant correlations with the way evaluators show awareness of what teachers have done to improve practice, encourage risk-taking, and invite teachers to develop their own improvement plan.

The 'take-away' from these correlations is how teachers perceive the conference to be more effective for strengthening teaching and holding teachers accountable when evaluators seek teachers' input for addressing concerns and issues by using formative approaches that encourage teachers to collaborate with them on actions to address concerns. As this correlation suggests, teachers are noting that evaluators who make an effort to work with teachers on issues and concerns, which is a formative approach to building collaborative and interpersonal relationships, contributes to the teachers' perceptions that the conference itself may be effective. This particular correlation has interesting implications for evaluators to use as a framework for developing protocols that should make the evaluation conference more effective for them and teachers. However, what is *not* indicated by this correlation is why those particular behaviors contribute to effectiveness over the other factors-items that address teachers' accomplishments and/or strengths, identify their concerns about the teachers' practice, and discuss anything related to compliance, rubric elements and/or reasoning for ratings given to the teacher.

A set of regression analyses of evaluators' specific conference approaches and general evaluation behaviors most strongly correlated with conference effectiveness provides a model that can be used to predict evaluation conference effectiveness. The regression models that account for the most variation in conference effectiveness measures (see Table 11: Composite Models Predicting Evaluation Conference Effectiveness) indicate the over-all conference effectiveness coalesces around evaluators who show encouragement through follow-up, exhibit fairness of appraisals, focus on student engagement, and exhibit an awareness of the teaching context and culture that impacts practice. The strongest predictors, focusing on these factors, possibly contribute to how the evaluation feedback conferences can strengthen evaluator-teacher partnerships and support teacher improvement, reflecting the formative purpose of evaluations over-all. The regression models also show that actions of evaluators which focus on summative purposes, such as promoting accountability, do not contribute as much to making the conference effective as the other behaviors. The data and regression analysis also suggest *how* evaluators employ those approaches when conducting the conference is connected to what makes the conferences effective for strengthening their teaching, building relationships, and holding teachers accountable. One implication of the factor correlations and regression models is that teachers' perceptions of effectiveness may be affected by the evaluators' general leadership style combined with what those evaluators do when conducting the conference itself. If the evaluators' over-all leadership style takes the formative approach, even when engaging in summative discussions, teachers note how the conference can still be deemed effective.

The over-all collection of survey responses reflects similar reactions as the teachers in the RAP interviews and then in the focus groups' q-sorts, but this interpretation of the survey data is limited by the objective reporting on evaluators' observed actions and behaviors. Even though

there are not enough total number of survey participants to make broader generalizations about the impact of evaluators' behaviors on conference effectiveness, the data do reveal distinctions between conference behaviors and general leadership approaches that are employed, simultaneously, in evaluation conferences that appear to predict the effectiveness of the conference. These complex correlations between and within evaluators' behaviors (the independent factor variables) with the conference effectiveness measures (the dependent factor variables) and the model of conference effectiveness require further research with a larger population before any generalizations can be drawn from these survey data.

Q-Sort Data

Q Methodology is a hybrid (MIXED) research approach that is unique for the way it opens up the investigation and interpretation of a phenomenon under study with both subjective and objective lenses. As a research method, the Q-approach employs a data collection instrument (i.e. scale-distribution of statements into a grid) that relies on the participants to draw from and report on their own subjective experiences with a specific phenomenon/topic. Once collected, the data from participants' grid distributions (i.e. q-sorts) are statistically analyzed for objective analysis while the participants' verbal/written responses about their *own* experience on the topic are qualitatively analyzed for subjective interpretation. Using the Q-approach, in addition to interviews and survey, capitalizes on the hybridity (QUAL-QUANT) of methods for data collection, which can be statistically analyzed, that allows for an additional means to clarify the perceptions of individual experiences across the three data sets. As the data from interviews and the survey reveal, this topic brings out multi-faceted and strong perceptions from all participants, and gathering both subjective and objective responses on such a topic using only a qualitative or a quantitative approach does not fully capture the complexity of the responses to their

experiences. Even though a Q-approach does enhance understanding the complexity of teachers' perspectives on this topic, the small number of participants in the Q-sample groups is a limitation to the generalizability of findings.

For this study, examining the collected q-sort distribution grids with this hybrid lens reveals similar experiences across districts and years of experience with evaluator behaviors that impact perceptions of conference effectiveness; over-all, the raw data show shared perspectives on objective evaluator behaviors which the participants consider most effective and useful for the evaluation conference to be considered effective. That being noted, the 13 participants statistically divide into two groups which 'define' specific behavioral factors (on the part of the evaluators) with an emphasis on certain approaches that impact the effectiveness of the evaluation conference. One group of nine participants gives precedence to whether evaluators are accurate, authentic, and fair in the way evaluations are conducted; the other group of four participants gives precedence to whether evaluators include useful feedback that shows awareness of what teachers are doing in the specific teaching context. The two perspectives are not contradictory, but they do emphasize different, as well as formative, aspects of how evaluators conduct conferences. In general, however, the perspectives of both groups focus on how the participants think *affectively* about what an evaluation conference reflects about *them* as teachers, especially when they perceive evaluators conduct an evaluation conference that focuses on formative collaboration and inter-personal connections.

A noteworthy dimension of evaluation conferences each group responded very strongly to, in terms of what is '**most ineffective/useful**,' is when they perceive evaluators conduct conferences that focus solely on APPR ratings or other summative topics. Both groups react negatively when they feel evaluators default to having summative discussions only as a way to

convey feedback or make an APPR appraisal. What the two groups seem to be sensitive to is how such discussions marginalize or even dismiss what they do as teachers, thus the evaluators do not show any investment in *them* as teachers.

Another over-lapping factor between the two groups emerges around contextual factors shared across and within districts. The participants' discussion and comments on the changes in APPR systems note how both teachers and evaluators are confined by the APPR changes because all districts must implement an evaluation process that will fulfill NYSED regulations and mandates. All the participants acknowledge how teachers are mandated by law to have a fair and equitable evaluation, and evaluations are a necessary part of the profession, yet each district's APPR system usually is affected by organizational culture, policy mandates and district politics unrelated to the over-all NYS system which further complicate the relationships between teachers and evaluators in those districts. Teachers cite how the lack of attention or acknowledgment by evaluators on/about those external contextual issues impacts how they perceive evaluators are able to help them improve practice when all the evaluators focus on is the summative purpose of the conference. The issue this external construct highlights, in a way that the interview and survey participants imply, is how the 'most effective/useful' approaches an evaluator can use should focus on formative purposes that are within the control of teachers, administrators and evaluators rather than on external mandates out of everyone's control.

The Q-approach used for this data set clarifies and distinguishes how teachers feel about what happens in evaluation conferences that impact them the most. The shared perspectives from the q-sorts on specific behavioral factors and contextual issues related to APPR evaluations provide clarification on how deeply feedback conference experiences impact teachers at all levels of teaching service, novice to most veteran, and across content areas. Even though the total

number of participants in the q-sort focus groups is not enough to generalize across the population of NYS teachers, there is validity and credibility of the data that suggests there is a connection between how evaluators approach conducting feedback conferences and to what extent teachers feel those behaviors and approaches make the feedback conference effective for impacting their practice.

Center of Tension: The AFFECTIVE Connection

As much as teachers seem to be suspect of all things having to do with evaluations and the APPR law, none of the participants across any of my research activities ever mention that evaluations should not be conducted. On the contrary, teachers often mention how evaluations could (and *should*) be an important way to improve their practice. This perception emerges as a theme based on teachers' reporting how a formative approach over summative impacts their impressions of evaluation conferences effectiveness. That being noted, the collected data also reveal teachers' complex and varied reactions to and perceptions of any feedback on practice, taken as formative or summative, that is communicated in evaluation conferences.

Across the three sets of data collected, an underlying theme emerges that shows teachers have complex and varied experiences that impact their very personal responses to and reception of feedback given in the context of evaluation conferences. Interpretations of the subjective data, specifically, propose teachers' reactions/responses and perceptions represent how teachers make personal affective connections between evaluation ratings and professional self-awareness-identity as a teacher. As defined in Chapter 1, "affective responses" are the subjective, intrinsic and internalized emotions, beliefs and attitudes teachers personally and individually have and/or hold that reflects the professional identity they hold of themselves and their work within that professional organization. When administrators assign efficacy ratings within the evaluation

context, teachers take it personally as well as professionally because teachers think of themselves and their practice as intertwined; it is how they identify as teachers. As extreme as this may seem, teachers do feel deeply about their practice and take it very personally when some other person makes judgments about the merit, worth, and value of their practice without the consideration that what is being judged is rooted very deeply in someone's professional soul. This personal-professional identity connection comes through much of the raw and visceral language in the more critical comments by interview and focus group participants. As one RAP interview participant noted, "This [teaching] is my life." Not all of the participants expressed this affective connection between the professional/personal identity and evaluation ratings, but for those teachers who do feel this connection, an evaluation that does not consider or acknowledge this connection equally with the performance ratings the evaluators assign is taken as a personal and professional affront to their sense of self.

This affective connection seems to be rooted in how teachers are internalizing APPR evaluations in such a way that the professional narrative a teacher constructs from self-reflection and self-appraisal comes into conflict with the external appraisal by administrators or evaluators. The internal-external dichotomy is creating tension between the teacher's side and the evaluator's side of the same professional narrative. According to feedback theory and studies (Ilgen, et al., 1979; Jawahar, 2010; Kinicki, et al., 2004; Kluger & DeNisi, 1998), the person receiving the external feedback will ultimately internalize the feedback given as a summative judgement or appraisal of their self-worth, whether or not the feedback/evaluation was formative or summative. Teachers in this study articulate how they struggle with the formative-summative tension brought on by internalizing their evaluations, whether or not they agree with or consider the evaluation as an authentic representation of their practice. This struggle highlights the

personal-professional affective connection that teachers make between their self-reflections on practice and the evaluators' appraisal of that practice. For many teachers, the affective-professional connection they feel is further complicated by how they perceive the APPR evaluations shape their professional life-story narratives; for some of the interview and focus group participants, having a sense of control over their professional narrative determines whether or not they perceive feedback conferences, and/or the evaluation process as a whole, to be effective for having an impact on their practice. .

Teachers' narratives about their evaluation conference experiences and reflection on practice, defined by both the qualitative and quantitative data, show how important establishing a sense of narrative authority over one's own teaching practice is for teachers to develop a sense of self. Teachers internalize evaluations as judgements of their worth and value as practitioners, and evaluators' appraisals have an impact on how teachers conceptualize what the evaluation says about their practice through a formative lens. Teachers use this formative lens about what they know and are able to do as a means to understand their daily experience while teaching, and "teachers filter all experience...through their personal practical knowledge, and express their knowledge of teaching in practice through their own narrative authority" (Olsen & Craig, 2001, pg. 667-668). Because the narrative version of knowledge construction is transactional and formative, authority comes from experience and is integral as each person both shapes his or her own knowledge and is shaped by the knowledge of others. Thus, narrative authority becomes the expression and enactment of a person's personal practical knowledge that develops as individuals learn to authorize meaning in relationship with others.

Having the narrative authority, which is inherently internal and affective, can be empowering for teachers and lead to acts of disruptive discourse that would counter the external

discourse of the APPR evaluation scores and commentary by evaluators. Teachers need to possess, as well as recognize their possession of, this narrative authority in the context of the evaluation process, and evaluators need to recognize that authority as well if there is to be a reshaping of the current APPR culture. The shift towards realizing a systematic reshaping of the APPR process cannot move forward unless teachers feel this control over their professional narrative lives and evaluators collaboratively conduct the evaluation process for predominately formative purposes rather than for summative purposes.

Over-all, the data points to the APPR evaluation process and how evaluators conduct the conference as part of this process creating tension between the institution and the teacher (Miller, 2005). What emerged from the subjective data are teachers' experiences and perceptions with APPR evaluations taking away their narrative authority over their professional understanding of self that comes from a formative self-evaluation/appraisal; the objective survey data reveal how teachers perceive the evaluators' approaches contribute to supporting or suppressing that narrative authority when the evaluation appraisal/ratings are used predominately as summative judgements. What the data imply is the need to shift the culture related to APPR evaluations more towards the formative purpose that addresses how teachers affectively internalize feedback on their practice rather than approaching the feedback from the summative position. The shift does require a re-shaping of protocols, approaches and strategies for conducting all parts of the evaluation process with emphasis on the teachers' self-awareness and acknowledgment of their affective-formative understanding of evaluations over-all.

Implications for Theory and Practice

The premise for the over-all design of this study is based on being pragmatic about understanding teachers' experiences with one part of the process, the evaluation feedback

conference, and explore whether or not the way evaluators conduct conferences could be contributing to the current tension felt by teachers and educational leadership when districts implement mandated APPR evaluations. As discussed in the literature review, there seems to be no a priori macro-theory that guides the implementation of evaluation systems, including the way to conduct evaluation conferences, but there are multiple micro-theories that do guide educational leadership, supervision and evaluation practices which encourage praxis for conducting/implementing effective evaluation processes using those theories. What the data from this study reveal are complex teacher responses to and perceptions of how evaluators conduct evaluation feedback conferences based on leadership practices encouraged by the theoretical and practical literature on educational supervision and evaluation. The data findings are a step toward shaping a theory for school leaders and teachers to consider when developing a framework for conducting conferences, while also providing actionable insights administrators and evaluators should consider using to address the current tension teachers and evaluators are experiencing with evaluation conferences.

Affective Evaluation Conference Theory

When I discussed my decision to use multiple philosophical assumptions and theoretical frameworks in this study, I noted that there is no single theory that guides educators toward a balanced approach to evaluation feedback conferences. The relevant macro-metatheoretical frameworks related to Naturalistic Inquiry, Critical Social Constructivism, and Theory of Subjectivity (Beuving & de Vries, 2015; Creswell, 2013; Kincheloe, 1993, 1997, 2006; Kincheloe & McLaren, 2000; Lincoln, 1997; Lincoln & Guba, 1985, 2000; Patton, 2015; Phillips, 1997; Schwandt, 2000; Stephenson, 2014; Tedlock, 2000; Vidich & Lyman, 2000) each contribute a facet or dimension toward a way of thinking about what an a priori theory could be.

The relevant micro-substantive theories of feedback (Ilgen, et al, 1979; Kinicki, et al., 2004; Kluger & DeNisi, 1996, 1998), educational performance evaluation, and performance feedback conferences (JCSEE, 2009; Popham, 2013; Scriven, 1995) provide a way to outline the inherent complexity of performance feedback within the context of a constructed experience framed by the macro-theories. Thus far, there has not been a research study conducted that has tested the validity of an a priori theory which combines these macro/micro-theories.

The practical literature from the field of clinical supervision (Danielson & McGreal, 2000; Danielson, 2015; Marzano & Toth, 2013; Marzano, et al., 2011), mentoring (Behrstock-Sherratt, et al., 2013; Costa & Garmston, 2002; Lipton, et al., 2003), and educational supervision/evaluation (DiPaola & Hoy, 2014; Frontier & Mielke, 2016; Glickman, 2002; JCSEE, 2009; Lipton & Wellman, 2013; Marshall, 2013) provide a path to praxis using a combination of macro/micro-theories as part of the over-all organizations' evaluation system. The general frameworks, considerations and protocols in this literature address various complexities inherent in school organizations' individual contexts connected to implementing evaluation systems that require adhering to mandated policies, yet the data from this study show there seems to be an inconsistency for how each school organization interprets and/or implements those frameworks, considerations and protocols for conducting effective evaluations.

When these theoretical assumptions and practical frameworks are combined with the collected data, a more nuanced theory starts to take shape that puts teachers' perspectives, collective knowledge and sense-making of evaluations at the forefront of understanding the effectiveness of feedback conferences. The data from this study seem to suggest that teachers' affective reception of and response to feedback during the evaluation conference, which is absent from the literature and research on educational performance evaluations, should take precedence

in conference protocols rather than strict use ‘effective approaches/strategies’ evaluators’ are trained to use to engage teachers. As the literature and research propose, the protocols for engaging teachers positions the evaluators as the actor in the process who does the engaging, employing approaches or strategies in such ways that would elicit teachers’ responses. Those approaches and/or strategies the literature emphasizes evaluators should employ have the potential to undermine how teachers accept feedback or evaluation of practice, especially if teachers perceive the use of those approaches/strategies ignore or undermine their ‘affective’ feelings toward the feedback or evaluation. What is absent from the current theories, literature and research is how to employ those noted approaches/strategies in ways that will acknowledge the ‘affective’ impact of those approaches on teachers and how they perceive the effectiveness of the evaluation over-all.

A theory of an affective connection between feedback and how teachers respond to the feedback is supported by the way each particular set of data provides a different facet of teachers’ perspectives, beliefs, and reflections on evaluation conference experiences (Quintelier, et al., 2018). To clarify such a theory requires research that will “stress the socially constructed nature of reality...and the situational constraints that shape the inquiry” (Denzin & Lincoln, 2000, p. 8). Toward that goal, a larger inquiry study with a revised survey instrument and wider range of qualitative participation (as either follow-up interviews or q-sort activities) that focuses on this connection will be able to confirm or dispute the extent this connection exists in a way that is statistically verifiable and reliable.

Praxis: Shift Towards Affective Feedback Conferences

In addition to exploring a theory of evaluation feedback conference effectiveness, the data from this study provides practical findings educational leaders can use if they, or their

districts, are in the position to revise current APPR processes within their control (i.e. negotiated items, external constructs). From what the data show, to make evaluations work for all members in an organization, educational leaders should focus on thinking systematically about how to judge, appraise and determine the efficacy of teachers with evaluation systems that “sustain productive, collegial working conditions that allow teachers to work collectively in an environment that supports learning for them and their students” (Darling-Hammond, 2013, p. 3). As the educational leadership literature and this study’s data suggest, to work ‘collectively’ includes building the capacity for trust between teachers, teachers and evaluators, and teachers and the APPR system such that it is built upon transparency and collaborative decision-making on the functions and purposes of teacher evaluations (Kimball, 2002).

Transparency about the function evaluations serve and collaboration between teachers and administrators should be the foundation for making a systematic shift in evaluation procedures that forefront the formative purpose. This shift includes collaboration on how to incorporate teacher-centered evaluation procedures, such as peer coaching, action research, and portfolios of practice, into an established process that would encourage formative discussions about practice while also addressing summative issues within those discussions. A systematic affective-formative approach that focuses on acknowledging teachers’ agency for evaluations, balancing functions/purposes of evaluations, and building trust could shift teachers’ and educational leaders’ over-all perceptions of evaluations as authentic, fair and equitable representations of teaching and learning.

Teacher Agency

Educational organizations, in general, are complex socially-constructed systems in which teachers and school leaders have developed interpersonal relationships and established cultural

norms based on shared understanding of teaching and learning. Meaning and sense-making is a social construct that happens in our everyday activities; “[p]eople act according to the meaning they impute in situations” and “these behavioural consequences or social formations have an impact on how people define new situations, on how they continue to think and act” (Beuving & de Vries, 2015, p. 32). Teachers’ narratives, especially about power structures and professional knowledge, are context-bound social constructions that emerge from experiences with administrators and evaluators as part of their performance reviews (Elbaz-Luwisch, 1997; Goodson, 1997).

As an evaluative activity, the feedback conference is subject to variations from external constructs associated with the evaluation system design and implementation, namely the social, political, and professional culture of the organization, that will impact how teachers and administrators make meaning before, during, and after the conference experience (DiPardo & Potter, 2003). Teachers and administrators “actively give shape and coherence to [the] experience” (Kegan, 1994, p.199), which can impact how teachers and administrators then use that understanding to examine, question, reflect on and revise perceptions of the impact the evaluation experience has on practice (Cranton & Taylor, 2012). This meaning-making shapes the reality that is a construction of the participants involved and usually is influenced by the social or cultural context of the situation (Kegan, 1994; Merriam & Bierema, 2014; Merriam, Caffarella, & Baumgartner, 2007); however, the participants (i.e. the teacher and evaluator) will shape and form meaning based on how each perceives the experience through social, political, and professional culture lenses (Cranton & Taylor, 2012; DiPardo & Potter, 2003; Kegan, 1994).

For performance evaluations to help teachers develop expertise and reflect on pedagogical decisions, teachers need the agency and authority to determine what and how to

change or improve their practice as part of the evaluation process and evaluative culture (Bandura, 1994, 2006; Schön, 1983, 1987). Teachers' investment, or buy-in, for the evaluation process comes from the support and cultivation of self-efficacy and agency, and both are needed for the over-all functioning of the evaluation system. If school organizations approach teacher evaluation from this sociological position, the system of evaluation will be addressing the needs of teachers on an affective, formative level (Bandura, 2006). Formative approaches to change teaching strategies that come from the teachers' self-appraisal and judgment of their performance have more impact on student achievement than decisions made by evaluators for summative purposes of evaluating practices for accountability (Gargani & Strong, 2014; Muñoz, et al., 2013; Taylor & Tyler, 2012). The first step toward making this shift would be to 're-think' how evaluators are trained for required 'certification' as per NYS Education Law mandates. The training should be based on discussions that include the voices and affective input from all stakeholders who would be impacted by such a shift. Districts and school organizations should take into consideration this practical-formative approach to balance the functions and purposes of evaluations in general, thereby relieving some of the tensions caused by the lack of collaborative understanding that seems to currently exist.

Formative-Summative Balance

Across the data collection, the common element or dimension of the conference that put almost all the participants on edge and creates the most tension is related to summative assessment and ratings. When evaluators approach the conference from this position, teachers do not respond well or positively. The interview participants outright dismiss any evaluator or evaluation when there is any indication that an evaluator is 'going by the book' or using a check-off list from the observation rubrics or APPR plan. Survey data show no correlation with

conference effectiveness measures or factor reliability when items addressed approaches used for summative purposes or topics. The q-sort participants all put the summative-rubric discussion approaches at the “least-to-less effective” side of the distribution grid. As one of the focus group participants notes on her q-sort, “The rubric does not measure/evaluate affective teaching which leads to effective teaching.” What is interesting about this dimension is how the participants also acknowledge and understand why some evaluators may approach conferences and giving feedback from this position, but they are in agreement that evaluators should have enough confidence in themselves and the teachers they evaluate to know when to use such an approach.

For the feedback conferences to have any formative or summative impact, evaluators and teachers should have collaborative, trusting relationships that will support an understanding of how evaluation feedback and ratings reflect either or both purposes (Calabrese, et al., 2004; JCSEE, 2009; Tuytens & Devos, 2017; Muñoz, et al., 2013; Wassermann, 2015). By doing so, teachers should perceive the feedback to be formative, authentic and valuable, which will then impact their reflectivity on practice (Ilgen, et al., 1979; Lortie, 1975). If teachers perceive feedback as formative, authentic and valuable, then the conference situation can be deemed effective for addressing the affective-professional needs of the teachers and the organization. As the data from the survey imply, the complicating factors and contextual situations presented by combining organizational performance assessments (summative purpose) with individual performance assessments (formative purpose) impact how teachers experience and understand the evaluation feedback conference for having an impact on their practice.

The literature on educational evaluation theory and standards make the point that evaluation systems should balance high-stakes decisions (i.e. employment or tenure recommendations) made from performance appraisals with how those decisions impact the

feelings of security (i.e. authenticity, equity and fairness) held by the employees being rated (JCSEE, 2009; McLaughlin, 1990; Scriven, 1995). The data from this study highlight the necessity for educational leadership to decide how the results of the APPR ratings will be used (i.e. towards fulfilling an accountability agenda, addressing improvement of teaching or both) and then communicate that decision, with transparency, to teachers and evaluators alike. Giving precedence to the evaluation's summative purpose undermines using the evaluation for serving a formative purpose and ignores the existence of the affective connection teachers feel. The guidance from NYSED on how districts make this distinction is vague, at best, so individual districts and educational organizations have the opportunity to craft processes and approaches toward clarifying the use of ratings for formative and summative purposes. By taking the steps to make the distinction of purpose as part of developing a more effective evaluation plan, which includes explicitly engaging teachers in formative and affective discussions as an effective way to hold teachers accountable for their practice, should contribute to possibly achieving balance and restoring/building capacity for trust in the evaluation process over-all.

Trust

A notable dimension of the evaluation conference, namely issues of power/control that impact trust between teachers and administrators, did not seem to be as much of an issue as I expected. When issues about power and control do come up in interviews, constructed responses in the survey, and q-sort/focus group discussions, participants seem to use a different affective lens to assess the situation. When power/control issues are brought up, participants also note other external constructs such as APPR law or policies that neither the evaluator nor teacher could change. Trust issues, when mentioned or noted, seem more entwined with how participants judged the evaluator's leadership style than just his/her conduct of the feedback conference.

The trust participants have in the way their administrators act in general seeps into the conference experience, but those relationships seem to be more complicated than what came out in any one of the data sets. The studies and research literature agree that the way evaluators give and teachers receive feedback is a key dimension for effective performance evaluation conferences. There needs to be an intricate balance of the sociological nature of appraising teaching performance with the culture and context of the school organization, and this balance requires understanding the complexities and “contraries” (McLaughlin, 1990, p. 412) from the variations in purpose and function across school organizational contexts. Educational leadership, evaluation and supervision literature attempts to provide the protocols and strategies to use that would cultivate these interpersonal skills, but there still requires a philosophical shift on the part of the administrators-evaluators toward developing these skills beyond using the protocols and strategies in the way they are trained to follow to be ‘certified’ to conduct evaluations. When administrators-evaluators are trained for conducting evaluations, that training should include purposeful and explicit cultivation of evaluators’ interpersonal skills that build trust between all the members in the organization and develop the evaluators’ responsiveness to teachers’ affective perception and understanding of evaluation processes and purposes.

From Limitations to Implications for Further Research

Impact of Sample Demographics

A challenge in this study was the limited demographics of participants across the three sample groups. At the time of designing the over-all study, the decision to gather only the most general demographics was meant to help keep the focus on exploring the over-all topic with as much anonymity as possible. The over-all collected demographics for this study primarily included years of teaching, experience with an evaluation conference, and school context (i.e.

rural-urban-suburban). Apart from the interviews and focus groups, the survey participants were afforded the most anonymity by virtue of the online survey format. The intention for not asking about much more specific demographic information (i.e. age, gender, and ethnicity of both teachers and evaluators) along with time teachers have worked with individual evaluators was to avoid having teachers not participate or give that information for fear of giving too much identifying information. Even so, the low response rate to the survey, even when only general demographics are collected, presents a consideration for how to collect more defining information that would contribute to interpreting data with more credibility.

As opposed to the anonymity of the survey, I had more specific demographic information on interview and focus group participants. I purposefully asked colleagues to be participants for the interviews and the focus groups so I could populate my two sample groups to address a variety of teaching experiences, number of evaluations, and school contexts. This decision also meant that my representation across the groups for other demographic characteristics were limited: mostly female (only 5 of the 20 interview/focus group participants are male), between the ages of 25 and 55, and all white. Aside from this information, I also was aware of how all 20 participants have very distinct and varied experiences with evaluations and different contexts of teacher-evaluator relationships; even though I did not report on or include this emic perspective when interpreting the findings, there is a possibility that this information may have affected the findings, which is a limitation for using qualitative research methods in general.

Embedded in the limited types of demographic information from all 59 participants (this number includes the 39 anonymous survey participants and the 20 interview/focus group participants) is the individual evaluation conference experiences they were asked to focus on in their responses. The interview participants reported on composite experiences that represented

their perceptions over-all; when a singular experience is noted, the participant focuses on how it exemplified the general conference experience. The survey participants were asked to report on the 'most recent' conference experience as a way to avoid asking participants to choose a conference experience that may not have been as 'fresh in their minds,' or may inadvertently (purposefully) be 'revised' when reporting, thus skewing data. The focus group participants were directed to reference their most 'recent' evaluation conferences when sorting behavior statements, but all focus group participants include commentary on their perspective based on how those experiences represent multiple experiences. The findings from all three data sets did show how participants have common/similar experiences despite the idiosyncrasies related to individual interpretations of varied experiences, but there is no way to verify or check reliability of these factors in light of that characteristic of the data. Further research of the evaluation conference experience must include gathering other demographics this study indicates are important to understanding professional relationships, and directions for all participants to limit reporting to one particular conference experience.

Generalizability of Findings

As with any mixed methods research study focusing on a specific phenomenon, there were some challenges related to data generalization across the research samples/population inherent in each research methodology paradigm. The traditional limitation associated with qualitative research is the generalizability of the findings and conclusions drawn from subjective data collected from a very small number of participants. As stand-alone qualitative data sets, the interviews and focus groups' subjective responses can only suggest what evaluators' behaviors and/or approaches teachers experience which impact their perceptions of conference effectiveness. The addition of the survey's quantitative data was meant to provide the statistical

support to overcome the limitation of the qualitative data but the number of survey responses, although derived from a randomly drawn sample, was so limited that they could not support confident generalizations either. A traditional limitation associated with quantitative research is the objectivity of data that can only imply that variables and correlations exist, not provide causation between variables, which impacts generalizability of the findings from just the survey.

The low number of total responses, across all three research methods, impacts the overall generalizability of any conclusions drawn from the data of such small samples of the total population. However, even with such a small total sample for both qualitative and quantitative data sets, teachers across all three data sets thought similar factors were important. The evidence of clear effect sizes of the quantitative survey data offset the concerns small sample sizes posed to the legitimation of the collective inferences and findings for each data set.

Even though the quantitative data showed relationships between the independent and dependent variables that reflect common considerations reflected in all three sets of research participants, the groups of participants showed a range of attitudes, from extremely critical to moderately positive, about those common factors/experiences, which makes it still more difficult to draw broader generalizations. Most of the participants in the interviews and focus groups project a critical tone within their narratives/reflective constructed responses, and the objective nature of the survey items/prompts possibly disguises a similar subjective perspective that may have been felt by the survey participants. The tone in the constructed survey responses, which is a smaller response group than the interviews and focus groups combined (only 11 of the over-all number of survey participants elected to contribute constructed responses at the end of the survey), are not as critical of the evaluators as the interview and focus group participants. The over-all tone of the survey responses is more positive and any critical reflection is projected on

other extenuating contexts (i.e. APPR regulations/mandates) rather than directly at evaluators. Yet, all three groups of participants stress roughly the same set of factors as important to the effectiveness of evaluation conferences.

The design of any research study using human subjects must take precautions for protecting the anonymity of the participants seriously, but that consideration also limits building in follow-up interviews or discussions with participants who presented the most variation in tone (i.e. most critical to most positive). I did not have a means to explore why participants have a more critical/positive tone or hold more critical/positive perspective of the way evaluations are being conducted in their particular district. Since there was no follow-up step for asking about subjective opinions built into the design of the survey portion of the study, I was unable to further investigate how those differences may have skewed the patterns of the factors and variables that did result from the survey data. These issues highlight one of the next steps for further research, specifically gathering both subjective and objective data from a much larger sample of the population. The next stage of research will include a larger sample size to empirically validate the findings on those factors which seem most important to teachers for how evaluators' approaches and behaviors impact conference effectiveness, as well as afford the opportunity for participants to volunteer being identified for follow-up interviews or discussions so that such diverging perspectives can be investigated for how demographic characteristics or specific organizational contexts impact those perspectives.

Perspectives on Conference Experiences

In the over-all exploration of this study, other basic features embedded in the NYS APPR process were not included as topics of discussion or survey items, such as the application of the HEDI scale, impact of student assessment, or prohibition to include other non-observable

evidence of teaching practice. The scope of the study limited the research focus to just teachers' perspectives, reflections, attitudes and responses to *evaluation conferences* and the *approaches* of evaluators when conducting those conferences. Limiting the focus to one experience with the process does not allow for considering how teachers feel the design and administration of the APPR system over-all affects their attitudes towards the evaluation conference; this limitation may have impacted how teachers responded to survey questions and distribution statements. As an implication for further research, the next iteration of the survey and follow-up interviews or focus groups should include a means for participants to report on experiences with other APPR features such as student performance measures, other evidence of practice, and over-all HEDI rating matrix, which are all part of the APPR law. These features may influence the reactions and responses to evaluation conferences, of teachers and evaluators, that could also provide insight on practical approaches for school organizations to use when developing plans to implement those required parts of the law into evaluation plans.

At the most relevant level, a feedback conference, as a part of the evaluation system process, includes two participants: the teacher and the administrator-evaluator. How the two participants interact and engage in this part of the process may determine how effective the conference is for both participants; however, a limitation of this study is that only teachers' perceptions on their experiences are reported, leaving out the administrator-evaluators' perceptions of the experiences or self-reflection on how evaluation conferences are conducted. An emerging theme across the data shows the interpersonal relationship between teachers and evaluators for how an evaluator conducts a conference impacts the process as a whole, especially when the evaluators have various levels of experience as evaluators, but the design of this study intentionally focused on teachers' perceptions to understand their side of the conference

experience which has been less studied thus far. An implication for further research is to study what teachers reported as important or effective approaches and how those perceptions align with the perceptions of evaluators, with various levels of experience, on those same practices and for what purposes those practices serve. The implication for practical application of studying how teachers' and evaluators' perceptions align, or do not align, is to use the information to help structure the evaluation process and activities, such as feedback conference, using the best strategies and approaches both teachers and evaluators agree on that would make the evaluation more collaborative, collegial, and effective rather than filled with tension.

As mentioned, this study specifically focused on gathering perspectives of teachers on the conference experience and their perceptions with how evaluators conducted those conferences; however, permission to ask teachers in order to gather data during the school year, and to use school district emails/listserves, necessitated asking *superintendents'* permission which may have impacted the rates of participation—from school districts and teachers. Every public-school district has a 'hierarchy' of administration and departments that must be navigated to secure permission to conduct research, from superintendents to building administrators then teachers' union leadership and finally teachers, which has layers of cultural, political, and sociological contexts related to organizational operations, not just APPR evaluations. What is unknown in this survey is whether or not administrative intentions may have come into tension with how teachers really feel about the evaluation process in their respective districts. What the participating superintendents may have thought was a positive way to highlight a positive culture of APPR in the district may not have been the same as how teachers think about the APPR process in the district. Since this survey was anonymous, there is no way to contact teachers to ask if their participation, or lack of participation, was about time of year, union discord with the

district, general apathy toward the topic, or some other contextual factor that influenced the participants not to respond at all. To conduct further research from the teachers' position warrants careful consideration for whom to contact besides the superintendents, such as approaching the teachers' unions (at the local, state, or national level) or inviting union leadership as partners in the research process as a way to encourage more collaboration between teachers and administration leadership. This approach would also be a practical way to help teachers engage district leaders in the evaluation discussions in a way that could address any tensions that may exist regarding how evaluations are conducted.

Even with acknowledging how this study invited superintendents as the means to contact teachers, there also may have been an issue with district leaderships' underlying reason *to participate/not participate* in the study in general. In one case, the district superintendent consented to the study, provided all the necessary contact information and distributed to the building according to Qualtrics, but zero participant responses came through Qualtrics. Some of the responses to initial invitations from superintendents indicate that they had no interest for teachers to participate in research on evaluation processes. No other reasons were given, but one conclusion could be that there are issues with the evaluation process in those districts whose leadership were unwilling to explore or expose. Perhaps districts whose leaders were confident that their teachers were generally positive about performance evaluation agreed to participate, whereas those districts that expected negative responses did not, skewing over-all responses. The study design does not include inviting superintendents to give reasons for or against participation, which could possibly explain the disconnection between eagerness of the superintendents and the low response numbers from individual buildings. Navigating through

such leadership hierarchy warrants careful consideration when constructing a revised survey instrument for further research with a larger sample of teachers across New York State.

Summary

The more recent attention on how teacher evaluations contribute to how school organizations fulfill state and federal accountability requirements/mandates also has drawn attention to how evaluations are conducted. There seems to be a perception that many school organizations' evaluation systems are flawed for how they incorporate and balance the dual functions/purposes of supervision and evaluation (Darling-Hammond, 2013, 2014; Frontier & Mielke, 2016; Hall, 2019; Popham, 2013) which then impedes the efficacy of evaluation systems to have the impact on teaching and learning that they are meant to have. To explore this issue, this study focused on three research questions:

1. What are teachers' experiences with how evaluators conduct evaluation feedback conferences?
2. Is there a connection between the way evaluators conduct feedback conferences, as reported by teachers, and how those conferences affect teachers' practice?
3. Based on teachers' reports, under what circumstances (if any) is it possible for evaluation feedback conferences to serve both summative and formative purposes?

The first question focused collecting data on what teachers report to be their lived experiences, collective knowledge and sense-making of the evaluation feedback conference experiences. The second question collected data on whether and how teachers perceive evaluators provide feedback during the conference that is useful and actionable. The third question provided a lens to investigate the possibility of the feedback conference being the key to reconciling the tensions participants feel in the evaluation process when the process serves two purposes.

The design of the study used three research methods to collect data addressing the research questions: naturalistic inquiry, survey research, and Q Methodology. On the qualitative

side of the study, interview narratives provided deeply personal and affective reactions anyone in the field has either felt or sensed as some time during a teaching career, yet the subjective and limited nature of qualitative studies brings out the issues of generalizability across larger populations and/or samples. On the quantitative side of the study, survey responses provided a means to gather larger sets of data from broader representation of participants than interviews, yet the objective nature of survey research limits the way the data can be interpreted and generalized across the population/sample if there is not enough participation, as happened in this study. The mix of subjective and objective approaches of Q Methodology for data collection, when used in tandem with the strictly qualitative and quantitative methods in this study, provided the corroborative data that clarified the interview narratives and survey reports which gave more validity and legitimacy to the conclusions drawn from all data.

Even though the three sets of data reveal similar factors that describe participants' evaluation conference experience for how evaluators conduct that conference, there are also differences (some very critical and others positive) in how participants perceived those experiences and evaluators' approaches for making the conference effective. On the qualitative side of the study, the interviews and q-sorts allowed participants to express opinions, reflections and attitudes (i.e. affective responses) which are critical of what they felt happening at the time but also reflect on positive experiences of evaluation. The quantitative data collected from the survey questionnaire highlight common factors across the small survey sample about the evaluator behaviors and professional context which contribute to the perceptions of evaluation conferences being effective; the generally positive tone of the survey participants' quantitative and qualitative responses seem to diverge from the interviews and q-sorts. All three groups of respondents, those who were critical of the evaluative feedback process and those who were

more positive, stressed the importance of the same set of factors: interpersonal trust, credible feedback, useful feedback, evaluators' mindfulness, and teachers' active engagement in the evaluation discussion. The small scope of this dissertation study, as one of its limitations, did not permit me to fully explore or make generalizations about specific dynamics of individual feedback conferences that would explain the divergent opinions or specifically identify approaches that contribute to the common factors. However, the tentative conclusions drawn from the three data sets and predictive strength of the models based on the survey factors and regression analysis contributes to an emerging theory, which a revised survey administered to a more representative sample of teachers would validate.

This research study is an initial stage of investigating a larger question that underlies many school organizations' evaluation systems: can school organizations and teachers resolve or reconcile issues and tensions related to the way school organizations conduct evaluative activities such as feedback conferences to address dual purposes simultaneously within the context of evaluation systems? The collected data do not give any definitive answers, but the data do indicate a potential over-arching theory of educational leadership specifically related to conducting feedback conferences is taking shape, which requires more research. Aside from generating evidence for a potential theory, the data collected do provide a practical path to addressing this question which school organizations could incorporate into current APPR plans. The data reveal a need for an over-all reshaping/shift in the culture around APPR systems, and school organizations could/should consider the findings as an opportunity to develop a framework for training evaluators to conduct effective feedback conferences with explicit emphasis on the formative use and purpose of evaluation ratings to address the affective connection teachers have with those evaluations and rating. Developing a framework for

evaluation conferences from the formative position will be effective for all participants and would (re)engage both teachers and administrators in evaluative endeavors that would promote powerful impacts on teaching and learning.

Appendix A:
Research Apprenticeship Interview Guide

Research Apprenticeship Interview Guide

- Title: Study of the Impact of Evaluation Conferences on Teachers' Professional Practice
- Method: 60 minute semi-structured interviews recorded for voice only
- 8 participants of various experience
- Question Guide:

-How long have you been a teacher?

-When was the most recent APPR evaluation conference you have had?

-If not yet, then when do you expect to have one?

-if you don't think you will have one this year, can you tell me some of your expectations when you do have one?

-can you describe how you developed your expectations?

Or: -if not in this district, then in what situation did you have an evaluation conference?

-Thinking of that evaluation conference, will you describe how it went?

-What are some of the reasons you think this happened?

-[depending on the response to the question] Can you tell me about some of the things that went (well—badly—neutrally)?

-Thinking about this experience, how did the evaluator approach the evaluation conference?

-Can you describe your evaluator's approach to the conference?

-What are some of the procedures you noticed your evaluator using?

-What were some of your reactions or feelings before the conference?

After the conference?

-In what ways, if any, did the conference focus on students or did the evaluator focus just on you?

-What do you think about the roles you and the evaluator have in APPR conferences?

-Can you describe how 'in control' you were during the conference?

-Do you feel the evaluators controlled all aspects or most aspects?

-Do you feel you and the evaluator controlled equal aspects of the conference?

-How were the responsibilities for the conference divided?

-Do you think the evaluation conference feedback has had an impact on what you do?

-What are some of the things you remember doing in class after the conference?

-How receptive do you feel you were to what the evaluator said?

(Can you elaborate on that?)

-What are some of the things you were open to [**or** not open to] to what the evaluator said?

- Can you remember any actions or suggestions of the evaluator that contributed to this feeling?
- What are some of the ways that you feel this evaluation conference fits in with the rest of the APPR evaluation system?
 - Can you tell me about other conversations you have had about the APPR system that has (or has not) had an impact on what you do in the classroom?
 - Is the evaluation conference useful to you?
 - What are some effects you have noticed in your practice that you feel resulted from these conversations?
- Do you think recent evaluation conferences have changed much from your experience in the past? What are some similarities/differences you notice?
- Do you have any other thoughts about APPR evaluation conferences that you feel you have not already included in our conversation?

Appendix B

Syracuse University Study Council Teacher Evaluation Survey, 1995

A Survey of Teachers and Other Professionals
in New York State Schools

**Performance Evaluation Systems
for School Professionals**

Sponsored by:

The Central New York School Study Council/
School Board Institute

and

The Syracuse University
School of Education
Teaching and Leadership Programs

Introduction

On the following pages, you will find questions about the procedures used to observe and evaluate your performance and the performance of other teachers and professionals in your school, as well as questions about conditions that some people have suggested might influence the effectiveness of evaluation systems.

Please read the questions carefully, and answer each as honestly and as frankly as possible. Most of the questions can be answered by circling a number 1 2 3 or 4. If you do not find the exact answer which fits your opinion, choose the one which comes closest to it. Feel free to write in any explanations or comments you may have anywhere in the margins or in the space provided at the end of the questionnaire.

If you are not a classroom teacher: The wording of several of the survey items refers specifically to the situations of classroom teachers. We are interested, however, in the observations and opinions of all professional staff members. If the wording of a survey item doesn't exactly fit your situation, please pick the answer that seems to fit your situation the best. If an item doesn't have any application to you at all, feel free to leave it blank.

Your answers to these questions will be strictly confidential. The analysis of the survey is being performed by university researchers. Only group responses will be reported, and your school will not be identified in any report of overall responses. Your own answers will be combined with those of other teachers, and used to prepare a description of the processes and contexts of performance evaluation in schools throughout New York State.

Thank you for your cooperation.

Observation and Evaluation Procedures

1. How accurate is each of the following statements in terms of your own current experience with observation and evaluation procedures in your school?

	Very accurate	Tends to be accurate	Tends to be inaccurate	Very inaccurate
Concerning formal observation procedures ...				
a. Those who formally evaluate your performance observe you in the classroom with students <u>at least once</u> a year before evaluating you.	1	2	3	4
b. Evaluators formally observe you <u>three or more times</u> a year.	1	2	3	4
c. Evaluators notify you ahead of time before formally observing you.	1	2	3	4
d. Evaluators ask you what they should look for before they observe you.	1	2	3	4
e. When evaluators observe you, they do so for a complete lesson or class.	1	2	3	4
f. Evaluators meet with you after visiting your class to discuss their observations.	1	2	3	4
g. Evaluators periodically "drop in" and informally observe you during the year.	1	2	3	4
Concerning formal evaluation procedures ...				
h. You receive a formal written evaluation at least once a year.	1	2	3	4
i. You understand the procedures and criteria that will be used to evaluate you.	1	2	3	4
j. You have a choice about what procedures will be used to evaluate you.	1	2	3	4
k. You have a say about what aspects of performance evaluators will focus on.	1	2	3	4
l. Evaluators use a rating scale to distinguish between levels of performance.	1	2	3	4
m. Evaluators prepare written narratives about some/all areas they evaluate.	1	2	3	4
n. You and other teachers evaluate each other as part of the evaluation process.	1	2	3	4
o. Evaluation procedures involve setting goals and assessing progress toward them.	1	2	3	4
p. You and evaluators carefully discuss your evaluation before it is final.	1	2	3	4
q. You and evaluators resolve any disagreements before an evaluation is final.	1	2	3	4
Concerning setting and monitoring improvement objectives ...				
r. Evaluators urge you to take the lead in setting objectives for improvement.	1	2	3	4
s. Evaluators decide what you should do to improve without consulting you.	1	2	3	4
t. When improvement objectives are set, the targets are challenging.	1	2	3	4
u. Evaluators help you get any training needed to meet improvement objectives.	1	2	3	4
v. You and evaluators discuss your efforts to improve throughout the year.	1	2	3	4

2. The next items concern what aspects of your performance evaluators focus on. In your experience, how much attention do your evaluators indicate that they pay to each of the following aspects?

	Substantial attention	Moderate attention	Little attention	No attention indicated
a. Whether your teaching reflects evaluator-preferred teaching strategies	1	2	3	4
b. How you adapt your teaching to the situations or needs of your students	1	2	3	4
c. How your individual lessons fit into a sequence, unit or longer-range plan	1	2	3	4
d. Your students' motivation and attitudes toward learning	1	2	3	4
e. What your students have achieved or are learning	1	2	3	4
f. How you monitor and assess your students' learning	1	2	3	4
g. Your classroom management, student conduct and discipline	1	2	3	4
h. Your classroom's appearance	1	2	3	4
i. How you counsel or work with individual students	1	2	3	4
j. How you work with parents	1	2	3	4
k. How you work with other teachers and staff members	1	2	3	4
l. How you conform to district/school policies	1	2	3	4
m. How you contribute to meeting district or school goals	1	2	3	4
n. Your efforts to improve your performance	1	2	3	4
o. Your efforts to innovate or develop new approaches to teaching	1	2	3	4
p. How your professional needs change as your career develops	1	2	3	4

3. In your experience, how much attention do evaluators indicate that they pay to each of the following sources of information when they prepare their evaluations of your performance?

	Substantial attention	Moderate attention	Little attention	No attention indicated
a. Their observations of your classroom performance	1	2	3	4
b. Your own assessments of your teaching	1	2	3	4
c. Your own assessment of your students' progress	1	2	3	4
d. A portfolio or other documentation you provide of your teaching	1	2	3	4
e. Your plans, materials, and other evidence of preparation for instruction	1	2	3	4
f. Your students' reports about your teaching	1	2	3	4
g. Parents' reports about your teaching	1	2	3	4
h. Your students' performance on standardized tests	1	2	3	4
i. Student demonstrations or other direct evidence of learning	1	2	3	4

Your Own Job and Responsibilities

1. This section concerns your own job and job responsibilities and which aspects you find particularly challenging. Please indicate how easy or difficult you find each of the following aspects of teaching. (If you don't perform the responsibility or confront the situation listed, please circle "DP/C" in the first column.)

	Do not perform or confront	Very easy	Somewhat easy	Somewhat difficult	Very difficult
a. Maintain classroom control	DP/C	1	2	3	4
b. Establish classroom routines	DP/C	1	2	3	4
c. Gain the respect of pupils	DP/C	1	2	3	4
d. Build rapport with students	DP/C	1	2	3	4
e. Keep records (e.g., grades, attendance, inventories, etc.)	DP/C	1	2	3	4
f. Apply what you learned about teaching in college preparation programs to your own teaching	DP/C	1	2	3	4
g. Prepare daily lesson plans	DP/C	1	2	3	4
h. Create your own instructional materials	DP/C	1	2	3	4
i. Choose appropriate strategies to teach particular content	DP/C	1	2	3	4
j. Identify the appropriate sequence for presenting material	DP/C	1	2	3	4
k. Decide how much content can be covered well in one lesson	DP/C	1	2	3	4
l. Anticipate and plan for problems that may arise in class	DP/C	1	2	3	4
m. "Read" classroom situations to identify when to change plans	DP/C	1	2	3	4
n. Address unexpected classroom situations while staying on task	DP/C	1	2	3	4
o. Evaluate the progress of individual students during class	DP/C	1	2	3	4
p. Respond to individual pupils' needs while teaching a group	DP/C	1	2	3	4
q. Identify student learning problems needing individual attention	DP/C	1	2	3	4
r. Adapt instruction to allow students to learn at different rates	DP/C	1	2	3	4
s. Control the pace of instruction.	DP/C	1	2	3	4
t. Lecture or make presentations to a whole class	DP/C	1	2	3	4
u. Conduct small group discussions in class	DP/C	1	2	3	4
v. Use exercises to teach students to solve problems on their own	DP/C	1	2	3	4
w. Use visual ways of presenting concepts	DP/C	1	2	3	4

	Do not perform or confront	Very easy	Somewhat easy	Somewhat difficult	Very difficult
x. Use computers in your instruction	DP/C	1	2	3	4
y. Use cooperative learning strategies in class	DP/C	1	2	3	4
z. Use exercises that require students to improvise and play roles	DP/C	1	2	3	4
aa. Incorporate music in your instruction	DP/C	1	2	3	4
bb. Address different student learning styles in the same lesson	DP/C	1	2	3	4
cc. Develop different student thinking skills in a single lesson	DP/C	1	2	3	4
dd. Integrate more than one subject in a single lesson	DP/C	1	2	3	4
ee. Give students choices of experiences to meet your objectives	DP/C	1	2	3	4
ff. Experiment with new ways of teaching or assessing students	DP/C	1	2	3	4
gg. Know when your students have learned what you want them to	DP/C	1	2	3	4
hh. Use student performance data to adjust your teaching	DP/C	1	2	3	4
ii. Evaluate your own teaching	DP/C	1	2	3	4
jj. Conduct research on teaching and/or student learning	DP/C	1	2	3	4
kk. Stay abreast of content knowledge changes in subject(s) you teach	DP/C	1	2	3	4
ll. Change your teaching significantly due to a new job assignment	DP/C	1	2	3	4
mm. Change your teaching significantly because of changes in curriculum, instruction or student assessment policies	DP/C	1	2	3	4
nn. Involve parents in their children's learning	DP/C	1	2	3	4
oo. Address parental concerns or complaints	DP/C	1	2	3	4
pp. Coordinate instructional approaches with other teachers	DP/C	1	2	3	4
qq. Co-teach the same group of students with another teacher	DP/C	1	2	3	4
rr. Work with specialists to address the needs of individual students	DP/C	1	2	3	4
ss. Work with colleagues to improve school programs	DP/C	1	2	3	4
tt. Serve as a mentor or coach for other teachers	DP/C	1	2	3	4
uu. Explain your reasons for instructional decisions to others	DP/C	1	2	3	4
vv. Conduct workshops, publish articles, or make similar professional contributions	DP/C	1	2	3	4
ww. Maintain your enthusiasm for teaching	DP/C	1	2	3	4
xx. Balance job responsibilities with personal needs/responsibilities	DP/C	1	2	3	4

The Instructional Environment

1. How accurate is each of the following statements about the instructional environment in which you work?

	Very accurate	Tends to be accurate	Tends to be inaccurate	Very inaccurate
In your school system ...				
a. You receive the instructional resources you need to perform your job.	1	2	3	4
b. The size of your classes is appropriate given your subject(s) and students.	1	2	3	4
c. Students are assigned to your classes according to their ability level.	1	2	3	4
d. You group students <u>within</u> your classes according to their ability level.	1	2	3	4
e. The needs of students with special learning problems or disabilities are addressed outside regular classrooms.	1	2	3	4
f. The public receives information on student performance in your subject(s).	1	2	3	4
g. Parents pay close attention to student performance in your subject(s).	1	2	3	4
h. The school board pays close attention to student performance in your subject(s).	1	2	3	4
i. The planning and revision of curriculum policies is an on-going process.	1	2	3	4
j. Information on student performance is used to evaluate curriculum policies.	1	2	3	4
k. You have a voice in the curriculum policies you are expected to follow.	1	2	3	4
l. New curriculum and instruction policies are adopted without assessing their impact on existing ones.	1	2	3	4
m. You receive training when new curriculum policies are adopted.	1	2	3	4
The curriculum and instruction policies you are expected to follow ...				
n. Reflect high expectations for all students.	1	2	3	4
o. Rely on student textbooks and other commercial materials.	1	2	3	4
p. Rely on instructional materials you have developed.	1	2	3	4
q. Rely on commercially-prepared lesson plans.	1	2	3	4
r. Expect you to follow specified teaching strategies.	1	2	3	4
s. Expect you to adapt your teaching to your students' progress.	1	2	3	4
t. Make use of short-answer tests to measure student learning.	1	2	3	4
u. Make use of student portfolios or other forms of authentic assessment.	1	2	3	4

2. How much emphasis is placed on each of the following considerations by the curriculum and instruction policies that you are expected to observe?

	Heavy emphasis	Moderate emphasis	Light emphasis	No emphasis at all
a. Ensuring coverage of specific content	1	2	3	4
b. Developing student mastery of certain basic skills	1	2	3	4
c. Promoting active participation of students in their own learning	1	2	3	4
d. Fostering student cooperation and learning from each other	1	2	3	4
e. Meeting State curriculum requirements	1	2	3	4
f. Fostering critical thinking and problem solving	1	2	3	4
g. Meeting the needs of individual students	1	2	3	4
h. Preparing students for standardized tests	1	2	3	4
i. Developing students' study skills	1	2	3	4

3. How accurate is each of the following statements about your relationships with other faculty members:

	Very accurate	Tends to be accurate	Tends to be inaccurate	Very inaccurate
a. You work with other teachers to develop instructional materials.	1	2	3	4
b. You work with other teachers on instructional objectives and strategies.	1	2	3	4
c. Specialists and classroom teachers work closely to address student needs.	1	2	3	4
d. When you have a problem or question, you ask other teachers for advice.	1	2	3	4
e. You and other teachers share information about how you teach.	1	2	3	4
f. When you try new things, you are reluctant to discuss them with others.	1	2	3	4
g. You try to avoid making commitments outside your classroom.	1	2	3	4
h. You and other teachers in your school agree on <u>what</u> students need to learn.	1	2	3	4
i. You and other teachers in your school agree on <u>how much</u> students can learn.	1	2	3	4
j. You and other teachers in your school agree on how you should treat students.	1	2	3	4
k. You and other teachers in your school are a team that works well together.	1	2	3	4
l. You and other teachers in your <u>department</u> or <u>grade level</u> are a team that works well together.	1	2	3	4

Relationships with Your Supervisor

1. This section deals with your relationships with the person you would identify as your primary supervisor.

Before responding to the items in this section, please review the following list and provide three kinds of information: (1) Circle the S beside the title of anyone who plays a role in supervising your performance (circle all that apply). (2) Circle the E beside the title of anyone who prepares a formal evaluation of your performance (circle all that apply). (3) Circle the PS beside the title of the one person you consider your primary supervisor (i.e., the person who most frequently supervises your teaching).

	Supervisor	Evaluator	Primary supervisor (circle <u>one</u>)
Central office curriculum or program coordinator	S	E	PS
Building principal	S	E	PS
Assistant principal or vice-principal	S	E	PS
Department or grade-level chair	S	E	PS
Other (please specify):	S	E	PS

Please indicate how accurate each of the following statements is about the one person you indicated is your primary supervisor:

	Very accurate	Tends to be accurate	Tends to be inaccurate	Very inaccurate
Your primary supervisor ...				
a. Demonstrates an understanding of the instructional problems you confront.	1	2	3	4
b. Demonstrates an understanding of the subject matter you teach.	1	2	3	4
c. Gives you helpful feedback about your teaching.	1	2	3	4
d. Minimizes classroom interruptions.	1	2	3	4
e. Gets you the information you need to perform your responsibilities.	1	2	3	4
f. Secures the resources you need to implement new ideas.	1	2	3	4
g. Gets you the professional development opportunities you need.	1	2	3	4
h. Communicates a clear vision of what your school/program should accomplish.	1	2	3	4
i. Communicates a clear sense of how students should be treated.	1	2	3	4
j. Communicates a clear sense of how staff members should treat each other.	1	2	3	4
k. Pays attention to your students' progress and outcomes.	1	2	3	4
l. Keeps his/her promises.	1	2	3	4
m. Explains if he/she is unable to keep a promise.	1	2	3	4
n. Accepts responsibility for his/her own decisions and actions.	1	2	3	4

	Very accurate	Tends to be accurate	Tends to be inaccurate	Very inaccurate
o. Shares ideas with you.	1	2	3	4
p. Shows appreciation for your work.	1	2	3	4
q. Gives others credit when good things occur in your school.	1	2	3	4
r. Takes credit when good things occur in your school.	1	2	3	4
s. Is open to criticism of his/her decisions.	1	2	3	4
t. Ignores advice he/she doesn't agree with.	1	2	3	4
u. Solicits advice from only a few teachers.	1	2	3	4
v. Consults you if he/she disagrees with your suggestions.	1	2	3	4
w. Involves you when important decisions that will affect you are made.	1	2	3	4
x. Makes sure that you understand the reasons for decisions that affect you.	1	2	3	4
y. Encourages you to take the lead on issues that concern you.	1	2	3	4
z. Looks for chances to give you assignments that you might find challenging.	1	2	3	4
aa. Encourages you and other teachers to turn to each other for help.	1	2	3	4
bb. Provides teachers with time to plan or discuss work together.	1	2	3	4
cc. Urges you to try new ideas, even if doing so might mean making mistakes.	1	2	3	4
dd. Avoids dealing with conflicts.	1	2	3	4
ee. Avoids making decisions unless they are absolutely necessary.	1	2	3	4
ff. Waits until problems arise before taking action.	1	2	3	4
gg. Acts quickly when problems arise.	1	2	3	4
hh. Encourages you to bring problems to his/her attention.	1	2	3	4
ii. Makes sure you and other teachers follow school policies.	1	2	3	4
jj. Holds you responsible if your students' achievement falls short of his/her expectations.	1	2	3	4
kk. Acknowledges when you have met agreed-upon objectives.	1	2	3	4
ll. Makes sure you understand what he/she expects of you.	1	2	3	4
mm. Encourages you to question conventional ways of doing things.	1	2	3	4
nn. Gives you confidence in your ability to do things that might seem impossible.	1	2	3	4

Professional Development

1. Based on your own experience, how effective has each of the following possible forms of professional development been in helping you develop the knowledge and skills you need to perform your responsibilities? (If you have not used or had access to a particular kind of development at all, circle "NA" in the first column.)

	Not available or used	Very effective	Somewhat effective	Somewhat ineffective	Very ineffective
a. Undergraduate education	NA	1	2	3	4
b. Graduate courses to complete certification	NA	1	2	3	4
c. Graduate courses pursued after certification	NA	1	2	3	4
d. <u>Required</u> district inservice training	NA	1	2	3	4
e. <u>Voluntary</u> district inservice training	NA	1	2	3	4
f. Professional development pursued outside your district	NA	1	2	3	4
g. Professional development in curriculum, instruction, or student assessment strategies	NA	1	2	3	4
h. Professional development in teacher supervision, mentoring, or peer assistance	NA	1	2	3	4
i. Professional development in self-assessment or reflective practice	NA	1	2	3	4
j. Professional development in teaming, group decision making, school improvement, or other forms of collaboration	NA	1	2	3	4
k. Your direct experience as a teacher	NA	1	2	3	4
l. Consultation with other teachers	NA	1	2	3	4
m. Classroom observation of other teachers	NA	1	2	3	4
n. Consultation with teachers trained as mentors, coaches, etc.	NA	1	2	3	4
o. Consultation with district or school specialists (e.g., curriculum specialists, psychologists, guidance counselors, etc.)	NA	1	2	3	4
p. Attendance at professional conferences	NA	1	2	3	4
q. Reading professional journals	NA	1	2	3	4
r. Personal research, conducted alone or with others	NA	1	2	3	4
s. Working with others to improve school policies or programs	NA	1	2	3	4
t. General consultation with administrators	NA	1	2	3	4
u. Your primary supervisor's observation and feedback	NA	1	2	3	4
v. Formal evaluations of your performance	NA	1	2	3	4

Overall Assessments of Observation and Evaluation Procedures

1. A school district's procedures for observing and evaluating performance can potentially serve a variety of different purposes. This section asks for your overall assessments of how effective your own district's observation and evaluation procedures are at serving each of the purposes listed below. (If you are unsure or have no basis for responding to a particular item, please circle "DK" in the first column.)

	Don't know	Very effective	Somewhat effective	Somewhat ineffective	Very ineffective
a. Helping new teachers learn basic skills	DK	1	2	3	4
b. Helping experienced teachers expand their knowledge and skills	DK	1	2	3	4
c. Providing teachers with useful feedback on their performance	DK	1	2	3	4
d. Providing recognition to teachers for exceptional efforts	DK	1	2	3	4
e. Providing for serious discussion of different approaches to teaching	DK	1	2	3	4
f. Encouraging teachers to take risks and try new things	DK	1	2	3	4
g. Ensuring that teachers conform to district/school policies	DK	1	2	3	4
h. Providing grounds for sound tenure decisions	DK	1	2	3	4
i. Providing assistance to teachers with serious teaching problems	DK	1	2	3	4
j. Documenting teachers whose performance is unsatisfactory	DK	1	2	3	4
k. Identifying school or program conditions that impede teaching	DK	1	2	3	4
l. Fostering cooperation between teachers and administrators	DK	1	2	3	4
m. Holding teachers accountable for the way they teach	DK	1	2	3	4
n. Holding teachers accountable for students' learning and progress	DK	1	2	3	4
o. Identifying professional development needs of individual teachers	DK	1	2	3	4
p. Identifying the need for professional development programs	DK	1	2	3	4
q. Encouraging teachers to take on-going responsibility for their professional development	DK	1	2	3	4
r. Actively involving teachers in assessing their own performance	DK	1	2	3	4
s. Improving teachers' instruction	DK	1	2	3	4

Background Information

The questions in this section will be used to compare the responses of different groups of respondents in other parts of this survey. *No responses will be reported that would identify individuals or small groups of individuals.*

1. How many years have you been a teacher or other school professional? _____ years

2. What is your current employment status (circle one):
 - a Probationary (tenure track)
 - b Tenured
 - c Temporary/contract employee

3. Which of the following describe your current position (circle any that apply):

a Regular subject classroom teacher	d Specialist (psychologist, counselor, librarian, etc.)
b Special education teacher	e Other (please explain):
c Special subject teacher	

4. Which of the following comes closest to indicating your level of education (circle one)?

a Bachelors degree or equivalent	d Graduate work beyond masters
b Some graduate training	e Graduate degree beyond masters
c Masters degree	f Other (please explain):

5. What is your gender: a Female b Male

6. Which of the following describe your race or ethnic background (circle any that apply)?

a African-American	d Hispanic-American
b Asian-American	e Native American/American Indian
c European-American/Caucasian	f Other

7. What is the gender of your primary supervisor? a Female b Male

8. Which of the following situations have you faced within the last three (3) years (circle all that apply)?
 - a You started working in this school district.
 - b Your job assignment has changed significantly in terms of duties, grade, or subjects taught.
 - c The curriculum you work with has changed significantly.
 - d You were assigned or started working with a new primary supervisor.
 - e You served as a department or grade level chair.
 - f You were active in efforts to revise your school's curriculum policies or programs.
 - g You were active in efforts to revise your school's teacher evaluation procedures.
 - h You got training or orientation in your teacher evaluation procedures.
 - i You received written information explaining your teacher evaluation procedures.
 - j None of the above apply.

9. One final question: If you could start your career over again, which of the following statements would come closest to describing the decision you would make (circle one)?
- a You would certainly choose to become a teacher or other school professional.
 - b You would probably choose to become a teacher or other school professional.
 - c Chances are about even that you would/would not become a teacher or other school professional.
 - d You would probably not choose to become a teacher or other school professional.
 - e You would certainly not choose to become a teacher or other school professional.

Additional Comments

We are interested in any additional comments you might have about the topics covered by this survey. Please use the back cover and/or additional pages if necessary.

Thank you for your participation.
Please return this survey in the accompanying envelope.

Appendix C

Teacher Evaluation Conference Survey, 2019



Teaching and Leadership Programs

**A Survey of Teachers' Experience with
Evaluation Feedback Conferences**

January 2019

Teacher Evaluation Conference Survey (2019)

Evaluation Feedback Conference Survey

We are interested in understanding the experience of teachers and other school professionals with evaluation feedback conferences. Our purpose is to improve the preparation of school administrators to conduct what is probably the most overlooked but arguably the most important part of the teacher evaluation process. We would like you to share your experiences from your most recent evaluation conference within the past two years. Please be assured that your responses will be anonymous and kept completely confidential. The study should take you around 15 minutes to complete. Your participation in this research is voluntary. You have the right to withdraw at any point during the study, or to decline to answer any particular question, for any reason, and without any prejudice or penalty. If you would like to contact the Principal Investigator in the study to discuss this research, please e-mail Dr. Joseph Shedd at jbshedd@syr.edu. By clicking the button below, you acknowledge that your participation in the study is voluntary, you are at least 18 years of age, you are a public school teacher or other school professional, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

- I consent, begin the study
- I do not consent, I do not wish to participate

The questions in this survey ask you discuss the most recent formal evaluation conference you participated in this year or last year. If the most recent conference was unusual or different in some respects, please focus on this conference but feel free to comment in the open-ended question space at the end of this survey on how it differed from other conferences you have had. If you are not a classroom teacher, feel free to skip over any questions that do not seem to apply to you. If you did not participate in any evaluation conference this year or last year, please indicate that in response to the next question; doing so will take you to the very end of the survey, where you can offer any comments you might want to make on your general experience with evaluation conferences.

Did you participate in at least one formal conference in which your own performance was evaluated and discussed in this or the previous school year?

- Yes
- No (skip to last open-ended question at the end of the survey)

Please indicate below the formal status of the person with whom you held your most recent evaluation conference. Please check below the one that best describes that person:

- Your building principal
- Another administrator in your building
- Another administrator who is your formal supervisor
- An “independent evaluator” who is not your formal supervisor
- A teacher or other colleague who serves as a “peer evaluator”
- Other (please describe): _____

The following is a list of specific topics that evaluators sometimes focus on when they conduct evaluation conferences with teachers. To what extent, if any, did the evaluator with whom you met in your most recent evaluation conference focus on each of the following:

	Quite a lot	A fair amount	To some extent	Not much	Not at all	Do not know/ Cannot recall
1. How you plan your lessons						
2. Your expectations for students						
3. Your knowledge of the content you teach						
4. Whether and how your students actively engage in learning						
5. How you assess students' learning						
6. Your classroom management and/or relationships with students						
7. How you adapt instruction for different students						
8. Your relationships with other teachers						

9. Your interactions with parents/care-givers/guardians						
10. Your plans for improving your teaching						
11. Student test scores						
12. Other data or information, besides test scores, on what students have learned						
13. Whether students are meeting expected learning standards or objectives						
14. Your compliance with district/school policies.						
15. The ratings the evaluator assigns you						

The following is a list of general approaches that evaluators sometimes use when they conduct evaluation conferences with teachers. How accurate is each of the statements about the approaches used by the evaluator with whom you met in your most recent evaluation conference:

	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Do not know/cannot recall
1. Let you know before the conference anything she/he wanted to discuss						
2. Explained what he/she wanted to accomplish in the conference						
3. Discussed all or most of the elements in the evaluation rubric your district uses						
4. Discussed a limited number of the elements in the rubric your district uses						

5. Discussed some or all of the ratings she/he planned to give you						
6. Explained her/his reasoning for the ratings she/he planned to give you						
7. Expressed appreciation for the work you do						
8. Discussed your general accomplishments and strengths						
9. Discussed concrete examples of your accomplishments or strengths						
10. Discussed ways of building on your accomplishments and strengths						
11. Discussed aspects of your teaching that he/she considered to be areas of concern						
12. Discussed concrete examples of anything she/he considered to be an area of concern						
13. Discussed what you would do to address areas of concern.						
14. Discussed which concerns or steps for improvement should receive highest priority						
15. Encouraged you to work with other						

teachers to address concerns that you and other teachers might share						
16. Made sure you and she/he agreed on what steps you would take as a result of the conference.						
17. Discussed specific things he/she would do to support your efforts to improve.						

The following is a list of things that evaluators sometimes do involve the person being evaluated when they conduct evaluation conferences. How accurate is each statement about the approaches used by the evaluator with whom you met in your most recent evaluation conference?

	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Do not know/cannot recall
1. Asked you before the formal observation what you thought he or she should look for						
2. Asked you before the conference to be prepared to discuss anything you thought needed attention						
3. Urged you during the conference to identify goals or concerns you wanted to discuss						
4. Encouraged you to provide information relevant						

to the topics being discussed						
5. Urged you to identify anything that might affect your ability to teach effectively						
6. Was open to your opinions, even if they might differ from her/his own						
7. Paid close attention to what you had to say						
8. Used strategies like paraphrasing, maintaining eye contact and other non-verbal cues to convey attention to your opinions						
9. Used open-ended questions that invited discussion rather than assertions that would close off discussion						
10. Encouraged you to suggest options for addressing identified concerns						
11. Invited you to choose among different options for addressing identified goals or concerns						
12. Emphasized the need for the two of you to reach conclusions you both could support.						

The following statements concern general situations and relationships that some people think might affect how evaluators and teachers conduct evaluation feedback conferences, even though they do not directly address the conferences themselves. Please indicate how accurately each of the following statements describes your experiences:

Teachers in your school...	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate
1. agree that all students can meet high expectations					
2. agree on what constitutes effective teaching					
3. regularly discuss curriculum and instructional issues with each other					
4. participate in ongoing professional development with each other					
5. coordinate their instruction with each other					
6. collaborate in developing and revising curriculum					
7. collaborate in developing common approaches to assessment of students					
8. work together to analyze data on student learning					

The administrator who most recently evaluated you...	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate
1. understands the curriculum you are responsible for following					

2. understands the instructional challenges you face					
3. observes you teaching on a regular basis					
4. asks for your advice on issues that affect you					
5. conveys a clear vision of what he/she wants your students and school to accomplish					
6. encourages teachers to turn to each other for advice					
7. encourages teachers to try new ideas, even if doing so might mean making mistakes					
8. takes the time to give each teacher's evaluation careful attention					

The items in this section ask you about the observations the evaluator made in your most recent evaluation conference. Please indicate how accurate each of the following statements is:

	Very accurate	Tends to be accurate	As accurate as not	Tends to be inaccurate	Very inaccurate	Do not know
1. The evaluator appraised your performance fairly						
2. The evaluator was respectful in how he/she shared opinions with you						
3. The evaluator used accurate information when discussing your performance						

4. The evaluator showed that she/he had the interests of your students in mind						
5. The evaluator showed an awareness of what you have done to improve your teaching.						
6. The evaluator showed that he/she is a good judge of your effectiveness as a teacher						
7. The evaluator helped you see situations you face from different vantage points						
8. The evaluator offered useful perspectives on things she/he observed in your teaching						
9. The evaluator made useful suggestions for specific things you might do to change your teaching						

The items in this section invite you to assess how effective you think your most recent evaluation conference was in terms of each of various criteria. How effective was that conference in terms of:

	Very effective	Fairly effective	As effective as not	Fairly ineffective	Not effective at all
1. Helping you improve your knowledge and skills					

2. Helping you make sense of problems or concerns you face in your teaching					
3. Helping you develop your own solutions for addressing identified goals or concerns					
4. Providing you with feedback that you could use to strengthen your teaching					
5. Providing recognition for your efforts					
6. Providing you with an opportunity to reflect on your own performance					
7. Providing an opportunity for serious discussions of different approaches to teaching					
8. Ensuring that you conform to district/school policies					
9. Identifying ways to further your professional development					
10. Arranging for you to get help or resources to improve your teaching					
11. Fostering trust between teachers and administrators					
12. Holding you accountable for the teaching strategies you use					
13. Holding you accountable for your students' learning					

Please provide any other comments about your experiences with the topics and/or interactions during evaluation feedback conferences not addressed above:
 (please limit response to 100 words)

This survey asked questions about your most recent evaluation conference. We understand that there are other factors that can have an impact on how conferences are conducted. If you feel that your most recent conference was unusual or different from previous experiences, or if you have

other observations you want to make about how evaluation feedback conferences are conducted in your school or district, please use the space below to explain:

How long have you served as a teacher or other professional staff member in your current school district?

- Less than one year
- One - three years
- Four - six years
- Seven - nine years
- Ten or more years

How long have you served as a teacher or other professional staff member in this and any combination of other school districts (indicate total for all districts)?

- Less than one year
- One - three years
- Four- six years
- Seven - nine years
- Ten or more years

Which of the following comes closest to describing your current assignment (check one)?

- Classroom, special education or special subject teacher
- Specialist primarily responsible for supporting teachers
- Other professional primarily providing support to students
- Other (please briefly describe):

Appendix D:
Dissertation Study IRB Materials

SYRACUSE UNIVERSITY



**INSTITUTIONAL REVIEW BOARD
MEMORANDUM**

TO: Joseph Shedd
DATE: April 13, 2018
SUBJECT: Exempt Protocol Review - Modifications Required
IRB #: 18-130
TITLE: *Conducting Teacher Evaluation Feedback Conferences*

The above referenced application, submitted for consideration as exempt from federal regulations as defined in 45 C.F.R. 46, has been evaluated by the Institutional Review Board (IRB) for the following:

1. determination that it falls within the one or more of the five exempt categories allowed by the organization;
2. determination that the research meets the organization's ethical standards.

It has been determined by the IRB that authorization of your protocol is deferred until you respond to the modifications required or issues raised below:

1. **Letters of cooperation must be obtained from each school district Superintendent and submitted to the IRB Office prior to conducting research in the school district. A sample letter template is available online or via the following link provided:**
<http://researchintegrity.syr.edu/wp-content/uploads/2018/01/Letter-Of-Cooperation-General-SAMPLE.doc-2018.doc>.
Please note that the researchers may submit letters from a school district for initial approval, and then add school districts via an IRB amendment request form once additional letters of cooperation are received.

Note: Electronic submission via email: orip@syr.edu is acceptable.

These required modifications should be addressed in a memorandum outlining changes; including highlighted changes to the application. Make sure to reference your IRB # on all communications. All correspondence should be sent to the address below within **ONE MONTH** of the date of this letter.

As a reminder, you may not initiate this human participants research project until the protocol receives IRB authorization.

Thank you for your cooperation in our shared efforts to assure that the rights and welfare of people participating in research are protected.

Tracy Cromp, M.S.W.
Director

DEPT: Teaching & Leadership, 150 Huntington Hall

STUDENT: Danielle Brain

SYRACUSE UNIVERSITY



**INSTITUTIONAL REVIEW BOARD
MEMORANDUM**

TO: Joseph Shedd
DATE: May 4, 2018
SUBJECT: Determination of Exemption from Regulations
IRB #: 18-130
TITLE: *Conducting Teacher Evaluation Feedback Conferences*

The above referenced application, submitted for consideration as exempt from federal regulations as defined in 45 C.F.R. 46, has been evaluated by the Institutional Review Board (IRB) for the following:

1. determination that it falls within the one or more of the five exempt categories allowed by the organization;
2. determination that the research meets the organization's ethical standards.

It has been determined by the IRB this protocol qualifies for exemption and has been assigned to categories 1 & 2. This authorization will remain active for a period of five years from May 3, 2018 until May 2, 2023.

CHANGES TO PROTOCOL: Proposed changes to this protocol during the period for which IRB authorization has already been given, cannot be initiated without additional IRB review. If there is a change in your research, you should notify the IRB immediately to determine whether your research protocol continues to qualify for exemption or if submission of an expedited or full board IRB protocol is required. Information about the University's human participants protection program can be found at: <http://orip.syr.edu/human-research/human-research-irb.html>. Protocol changes are requested on an amendment application available on the IRB web site; please reference your IRB number and attach any documents that are being amended.

STUDY COMPLETION: Study completion is when all research activities are complete or when a study is closed to enrollment and only data analysis remains on data that have been de-identified. A Study Closure Form should be completed and submitted to the IRB for review ([Study Closure Form](#)).

Thank you for your cooperation in our shared efforts to assure that the rights and welfare of people participating in research are protected.

Tracy Crump, M.S.W.
Director

DEPT: Teaching and Leadership, 150 Huntington Hall

STUDENT: Danielle Brain



SYRACUSE UNIVERSITY

SCHOOL OF EDUCATION

TEACHING AND LEADERSHIP PROGRAMS

150 Huntington Hall • Syracuse NY 13244 • 315-443-2685

A Study about Conducting Teacher Evaluation Feedback Conferences

Dear [superintendent]:

For all the attention and resources devoted recently to teacher evaluation, little attention has been paid to what many administrators say is the most challenging and important part of the evaluation process: engaging individual teachers in discussions of their performance. I am writing to seek your help in conducting a survey study of the evaluation feedback process.

[School name] is one of 100 New York schools that have been randomly selected from a list of all public schools in the state to participate in the study. The survey asks teachers to describe how their most recent evaluation conference was conducted and how effective they found the conference to be. The overall purpose of the study is to pinpoint how administrators might enhance the effectiveness of evaluation conferences, and recoup their districts' investments in the evaluation process.

Once the statewide survey data are analyzed, we would send you and your teachers' union a report summarizing the study's overall findings, but this report would not identify your district, school or administrators. The responses of individual teachers would be anonymous. If you and your union jointly request, we could provide you with a second report, comparing your own teachers' responses with those of teachers across the state; we would not share this report with anyone but you and your union president, although you could share it with others if you chose to do so. If you would like to arrange a time for me to meet with you, your administrators, and your union to discuss either of these reports and their possible use in identifying ways of improving the evaluation feedback process, I would be happy to arrange such a meeting.

If you decide to participate, we will ask you to distribute a brief description of the fifteen-minute survey and its purpose to all faculty members in [school name], and then to forward them an email request to fill out the survey. The request would include a link for them to use to access the survey. The survey would include a school code that we would use to aggregate each school's responses and calculate response rates, but it would not ask them to identify themselves or the administrator who conducted the evaluation conference.

If you have any questions about the study, please contact me at jshedd@svr.edu or at the phone or mail address listed above. If you decide to participate in this study, we ask that you copy the text of the attached cooperation letter onto your letterhead and return it to us by [date]. We are acutely aware of how challenging the evaluation process is for our state's administrators and teachers, and are committed to helping you and your colleagues find ways of strengthening that process.

Very truly yours,

Joseph B. Shedd, PhD
Associate Professor of Educational Leadership
Syracuse University School of Education

Invitation to Superintendents



SCHOOL OF EDUCATION
150 Huntington Hall ▪ Syracuse NY 13244 ▪ 315-443-2685

A Study about Teacher Evaluation Feedback Conferences

Dear [School] Teachers and Other Professional Staff Members:

With the permission of your superintendent and building principal, we are writing to invite you to participate in a brief online survey study being conducted by Syracuse University. The survey is designed to give you and other teachers and professionals in [School name] and a randomly-selected group of other schools across New York State a voice in identifying what enhances or limits the effectiveness of evaluation feedback conferences. In a few days, you will be receiving an email invitation to take the survey, with a URL link that will take you directly to the online survey.

The survey will ask you to indicate the accuracy of various statements about your most recent evaluation conference and how effective you found the conference to be. It will not ask you to identify yourself or the name of the evaluator who conducted the conference you are describing. Once you and teachers in other schools have completed the survey, we will send your district and its teachers' and administrators' unions a report summarizing the study's statewide findings, and if they wish, a separate report summarizing responses for teachers in your school. Neither report will identify you, your district, school or administrators by name. Your own responses will be confidential and anonymous. The survey should take between ten and fifteen minutes to complete.

Your participation in the study is entirely voluntary. You may choose not to participate, to stop answering the survey at any point, or to skip over any question you do not choose to answer, without any penalty. In fact, no one will know whether you will have answered the survey or not. Although there is always the risk of compromising privacy, confidentiality, and/or anonymity when researchers use email or the internet, your confidentiality will be maintained to the degree permitted by the technology being used. It is important for you to understand that no guarantee can be made regarding the interception of data sent via the internet by third parties.

If you have any questions about the study, please contact me at jbshedd@syr.edu or at the phone or mail address listed above.

Very truly yours,

A handwritten signature in blue ink that reads "Joseph B. Shedd".

Joseph B. Shedd, PhD
Associate Professor of Educational Leadership
Syracuse University School of Education

Invitation to Participate to Teachers



TEACHING AND LEADERSHIP PROGRAMS
150 Huntington Hall • Syracuse NY 13244 315-443-2685

A Study about Conducting Teacher Evaluation Feedback Conferences

Dear Teachers and Other School Professionals:

You recently received a letter about your district participating in a survey research study being conducted by Syracuse University School of Education. The survey invites you and other teachers and professionals in your school, and in a randomly-selected group of other schools across New York State, to help identify what enhances or limits the effectiveness of evaluation feedback conferences.

To conduct this survey study, the researchers are using an online software program, Qualtrics, to distribute and collect survey data. This program adheres to ethical guidelines and requirements for human research studies, and makes it possible for us to guarantee that your responses to the survey will be anonymous and confidential. As we explained in our previous letter, participation in the study is entirely voluntary. You can participate or not, start and then stop completing the survey, or decline to answer any particular survey question, at any time and without penalty. It should take between 10-15 minutes to complete the survey.

Qualtrics allows you to complete the survey using any device that accesses your email accounts (i.e. laptops, school desktop computers, smartphones) while maintaining confidentiality and anonymity.

To access the survey, please click on the following link, or cut-and-paste it into any browser:

[URL]

If you have any questions or concerns regarding the survey link, please contact me at jbshedd@syr.edu or at the School of Education number above.

Very truly yours,

A handwritten signature in black ink, appearing to read "JB Shedd".

Joseph B. Shedd, PhD

Associate Professor of Educational Leadership
Syracuse University School of Education

Initial Email Invitation to Teachers



TEACHING AND LEADERSHIP PROGRAMS
150 Huntington Hall • Syracuse NY 13244 315-443-2685

A Study about Conducting Teacher Evaluation Feedback Conferences

Dear Teachers and Other Professional Staff,

You recently received a letter about your district participating in a survey research study being conducted by Syracuse University School of Education, with a unique URL link to the survey. As you may recall, the survey is designed to give you and other teachers and school professionals in your school, and in a randomly selected number of other schools across New York State, an opportunity to help identify what enhances or limits the effectiveness of evaluation feedback conferences.

Since your participation is confidential, we do not know who has and has not responded to the survey. If you have already responded, the researchers would like to thank you. If you have not already completed the survey, please try to respond by [day, date]. The following is the URL link to the survey:

[URL]

If you have any questions or concerns regarding the survey link, please contact me at jbshedd@syr.edu or at the School of Education number above.

Very truly yours,

A handwritten signature in black ink, appearing to read "J. Shedd".

Joseph B. Shedd, PhD

Associate Professor of Educational Leadership
Syracuse University School of Education
Principal Investigator

Follow-up memo with URL to Teachers



SYRACUSE UNIVERSITY

SCHOOL OF EDUCATION
TEACHING AND LEADERSHIP PROGRAMS
150 Huntington Hall • Syracuse NY 13244 • 315-443-2685

A Study about Conducting Teacher Evaluation Feedback Conferences

Dear [Teacher]:

I would like to ask you to consider participating in a Focus Group to give your thoughts on data collected for my dissertation research on evaluation feedback conferences. My interest in this research topic stems from all the attention that has been devoted recently to how test scores, student learning objectives, and ratings are used in the APPR process, with little attention to teachers' experiences with evaluation conferences. Only a handful of researchers have asked teachers themselves what they think about these conferences. As part of the research process, I am convening a focus group of ten teachers to review the data findings and interpretations related to the findings.

The purpose and over-all design of my research study is to understand teachers' experiences with evaluation feedback conferences, and this focus group is the culminating portion of my research study. The focus group will look over survey data collected from randomly selected schools across New York State. The survey asked teachers to indicate the accuracy of various statements about their most recent evaluation conference and how effective they found the conference to be. All teacher responses were anonymous, and their participation in the study was entirely voluntary.

The focus group discussion should take only an hour of your time and will be voice recorded for research purposes. Your comments may be included in chapters of my dissertation as affirming or disputing my research findings, but your identity will be anonymous. As compensation for participating, you have the choice to receive a certificate of participation for one hour of CTLE-Pedagogy Professional Development for Professional Licensure requirements.

If you are willing to participate, please contact me at drbrain@sy.edu with your contact information (preferably your personal contact information unrelated to your school/district). Once I have 10 participating teachers, I will contact you to arrange the date, time and place for the focus group to meet. If you have any questions about the study, please contact me at drbrain@sy.edu, or my dissertation chairperson, Dr. Joseph B. Shedd at jbshedd@sy.edu.

Very truly yours,

Danielle R. Brain
Doctoral Graduate Student
Syracuse University School of Education

Focus Group Invitation to Teachers

References

- Alreck, P. & Settle, R.B. (1995). Composing questions. *The Survey research handbook, 2nd Ed.* Boston, MA: Irwin/McGraw-Hill. ISBN10: 0-256-20321-6.
- Arneson, S. (2015). Improving teaching, one conversation at a time. *Educational Leadership, 72(7)*: 32-36.
- Babbie, E. (1990). *Survey research methods, 2nd Edition.* Belmont, CA: Wadsworth Publishing Company.
- Baker, B.D., Oluwole, J.O., & Green, P.C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives, 21(5)*. EPAA/AAPE's Special Issue on *Value-Added: What America's Policymakers Need to Know and Understand*, Guest Edited by Dr. Audrey Amrein-Beardsley and Assistant Editors Dr. Clarin Collins, Dr. Sarah Polasky, and Ed Sloat. Retrieved 1 January 2019 from <http://epaa.asu.edu/ojs/article/view/1298>.
- Behrstock-Sherratt, E., Rizzolo, A., Laine, S., & Friedman, W. (2013). *Everyone at the table: Engaging teachers in evaluation reform.* San Francisco, CA: Jossey-Bass.
- Bandura, A. (2004). Swimming against the mainstream: The early years from chilly tributary to transformative mainstream. *Behaviour Research and Therapy, 42*: 613-630.
- Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1, 2*: 164-180.
- Benjamin, B. E., Yeager, A., & Simon, A. (2012). *Conversation transformation: Recognize and overcome the 6 most destructive communication patterns.* New York, NY: McGraw-Hill.
- Beuving, J. & de Vries, G. (2015). *Doing qualitative research: The craft of naturalistic inquiry.* Amsterdam, Netherlands: Amsterdam University Press.
- Bogdan, R.C. & Biklen, S.K. (2007). *Qualitative research for education: An introduction to theories and methods, 5th Ed.* Boston, MA: Pearson.
- Brookhart, S. M., & Moss, C. M. (2015). How to give professional feedback. *Educational Leadership, 72(7)*: 24-30.
- Calabrese, R.L., Sherwood, K., Fast, J., & Womack, C. (2004). Teachers' and principals' perceptions of the summative evaluation conference: an examination of model I theories-in-use. *The International Journal of Educational Management, 18(2)*: 109-117. DOI: 10.1108/09513540410522243.

- Cameron, R. (2011). Mixed methods research: The five Ps framework. *The Electronic Journal of Business Research Methods*, 9(2): 96-108. Available online www.ejbrm.com.
- Chambers, J.H. (1992). *Empiricist research on teaching: A philosophical and practical critique of its scientific pretensions*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Champ, C. H. (2015). Measuring teacher effectiveness: The impact of institutional culture on initial implementation of New York's annual professional performance review. *AASA Journal of Scholarship and Practice*, 12(2): 34-43.
- Charalambous, Y.C., Komitis, A., Papacharalambous, M., & Stefanou, A. (2014). Using generic and content-specific teaching practices in teacher evaluation: An exploratory study of teachers' perceptions. *Teacher and Teacher Education*, 41: 22-33. DOI: <http://dx.doi.org/10.1016/j.tate.2014.03.001>.
- Chase, S.E. (2005). Chapter 25-Narrative inquiry: Multiple lenses, approaches, voices. In *Handbook of Qualitative Research*, 3rd Ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Cherasaro, T.L., Brodersen, R.M., Yanoski, D.C., Welp, L.C., & Reale, M.L. (2015). *The Examining Evaluator Feedback Survey* (REL 2016-100). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Combs, J. P., Harris, S., & Edmonson, S. (2015). Four essential practices for building trust. *Educational Leadership*, 72(7): 18-22.
- Conley, S. & Glasman, N.S. (2008). Fear, the school organization, and teacher evaluation. *Educational Policy*, 22(1): 63-85. DOI: 10.1177/0895904807311297.
- Cooper, B.S., Ehrensals, P.A., & Bromme, M. (2005). School-level politics and professional development: Traps in evaluating the quality of practicing teachers. *Educational Policy*, 19(1): 112-125. DOI: 10.1177/0895904804272231.
- Costa, A.L. & Garmston, R.J. (2002). *Cognitive coaching: A foundation for renaissance schools*, 2nd Ed. Norwood, MA: Christopher-Gordon Publishers, Inc.
- Craig, C. J. (2007). Story constellations: A narrative approach to contextualizing teachers' knowledge of school reform. *Teaching and Teacher Education*, 23: 173-188.
- Cranton, P. & Taylor, E.W. (2012). Transformative learning theory: Seeking a more unified theory. In *The handbook of transformative learning: Theory, research, and practice*. Edited by Edward W. Taylor, Patricia Cranton and Associates. San Francisco, CA: Jossey-Bass.

- Creswell, J.W. (2013). *Qualitative inquiry & research design: Choosing among five approaches*, 3rd Ed. Thousand Oaks, CA: SAGE Publications, Inc.
- Creswell, J.W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: SAGE Publications, Inc.
- Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership*, 58(5): 12-15.
- Danielson, C. (2009). *Talk about teaching: Leading professional conversations*. Thousand Oaks, CA: Corwin Press.
- Danielson, C. (2015). Framing discussions about teaching. *Educational Leadership*, 72(7):38-41.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation: To enhance professional practice*. Princeton, NJ: Educational Testing Service.
- Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. In *The New Handbook of Teacher Evaluation*, Jason Millman & Linda Darling-Hammond, Eds. SAGE Publications Ltd.: 18-32. DOI: <http://dx.doi.org/10.4135/9781412986250>
- Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York, NY: Teachers College Press.
- Darling-Hammond, L. (2014). One piece of the whole: Teacher Evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1): 4-13, 44.
- Denzin, N.K. & Lincoln, Y.S. (2000). Introduction: The discipline and practice of qualitative research. *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Dillman, D., Smyth, J., & Christian, L.M. (2014). Chapter 2: Reducing people's reluctance to respond to surveys. *Internet, phone, mail, and mixed-mode surveys, 4th Edition*. Hoboken, NJ: John Wiley & Sons.
- DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20). Available online: <http://pareonline.net/getvn.asp?v=14&n=20>.
- DiPaola, M.F. & Hoy, W.K. (2014). *Improving instruction through supervision, evaluation, and professional development*. Charlotte, North Carolina: Information Age Publishing, Inc.
- DiPardo, A. & Potter, C. (2003). Beyond cognition: A Vygotskian perspective on emotionality and teacher's professional lives. In *Vygotsky's educational theory in cultural context*. Edited by Alex Kozulin, Boris Gindis, Vladimir S. Ageyev, and Suzanne M. Miller. New York, NY: Cambridge University Press.

- Donaldson, M.L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, 49(5): 838-882. DOI: 10.1177/0013161X13485961.
- Donaldson, M.L. (2016). Teacher evaluation reform: Focus, feedback, and fear. *Educational Leadership*, 73(8): 72-76.
- Donaldson, M.L. & Donaldson, G.A. (2012). Strengthening teacher evaluation: What district leaders can do. *Educational Leadership*, 69(8): 78-82.
- Drago-Severson, E. & Blum-DeStefano, J. (2014). Tell me so I can hear: A developmental approach to feedback and collaboration. *JSD: The Learning Forward Journal*, 35(6): 16-22.
- Dziopa, F. & Ahern, K. (2011). A systematic literature review of the applications of Q-Technique and its methodology. *Methodology*, 7(2): 39-55. DOI: 10.1027/1614-2241/a000021.
- Eden, S., Donaldson, A., & Walker, G. (2005). Structuring subjectivities? Using Q methodology in human geography. *Area* (a Royal Geographical Society with IBG publication), 37(4): 413-422. Stable URL: <https://www.jstor.org/stable/20004480>.
- Ellingsen, I.T., Størksen, I., & Stevens, P. (2010). Q methodology in social work research. *International Journal of Social Research Methodology*, 13(5): 395-409. DOI: 10.1080/13645570903368286.
- Fine, M., Weis, L., Weseen, S., & Wong, L. (2000). For whom? Qualitative research representations and social responsibilities. In *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Finster, M. & Milanowski, A. (2018). Teacher perceptions of a new performance evaluation system and their influence on practice: A within- and between-school level analysis. *Education Policy Analysis Archives*, 26(41): 1-48. DOI: <http://dx.doi.org/10.14507/epaa.26.3500>.
- Fowler, F.J. (1995). Some general rules for designing good survey instruments. *Improving survey questions, design, and evaluation*. Thousand Oaks, CA: SAGE Publications, Inc.
- Fowler, F.J. (2014). *Survey research methods*, 5th ed. Thousand Oaks, CA: SAGE Publications, Inc.

- Fraser, H. & Jarldorn, M. (2015). Narrative research and resistance: A cautionary tale. In Susan Strega & Leslie Brown, eds., *Research as resistance: Revisiting critical, indigenous, and anti-oppressive approaches*, 2nd Ed. Toronto, Ontario: Canadian Scholars' Press Inc.
- Freire, P. (1998). *Pedagogy of freedom: Ethics, democracy, and civic courage*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Frontier, T. & Mielke, P. (2016). *Making teachers better not bitter: Balancing evaluation, supervision, and reflection for professional growth*. Alexandria, VA: ASCD.
- Glickman, C. D. (2002). *Leadership for learning: How to help teachers succeed*. Alexandria, VA: ASCD.
- Gubrium, J.F. & Holstein, J.A. (1997). Chapter 1-Method talk. *The new language of qualitative methods*. New York, NY: Oxford University Press.
- Gubrium, J.F. & Holstein, J.A. (2000). Analyzing interpretive practice. In *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Hall, P. (2019). The instructional leader's most difficult job. *Educational Leadership*, 76(6): 12-17.
- Hallinger, P., Heck, R.H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26: 5-28. DOI: 10.1007/s11092-013-9179-5.,
- Halverson, R.R. & Clifford, M.A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Educational Administration Quarterly*, 42(4): 578-619. DOI: 10.1177/001316X05285986.
- Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York, NY: Teachers College Press.
- Heath, S.B. & Street, B.V. (2008). *On ethnography: Approaches to language and literacy research*. New York, NY: Teachers College Press.
- Hedrick, T.E., Bickman, L., & Rog, D.J. (1993). *Applied research design: A practical guide*. Newbury Park, CA: SAGE Publications, Inc.
- Helm, V.M. & St. Maurice, H. (2006). Conducting a successful evaluation conference. In *Evaluating Teaching*, James H. Stronge, Ed. Sage Publications, Inc. DOI: <http://dx.doi.org/10.4135/9781412990202>.

- Herlihy, C., Karger, E., Pollard, C., Hill, H.C., Kraft, M.A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, 116:1-28.
- Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2): 371-401.
- Hinchey, P.H. (2010). *Getting teacher assessment right: What policymakers can learn from research*. Boulder, CO: National Education Policy Center. Retrieved 1 January 2019 from <http://nepc.colorado.edu/publication/getting-teacher-assessment-right>.
- Hopkins, P. (2016) Teacher voice: How teachers perceive evaluations and how leaders can use this knowledge to help teachers grow professionally. *NASSP Bulletin*, 100(1): 5-25. DOI:10.1177/0192636516670771.
- Howard, B.B. & Sanders, J.R. (2006). Applying the personnel evaluation standards to teacher evaluation. In *Evaluating Teaching*, James H. Stronge, Ed. SAGE Publications, Inc.: 55-68. DOI: <http://dx.doi.org/10.4135/9781412990202>.
- Huber, S.G. & Skedsmo, G. (2016). Teacher evaluation—accountability and improving teaching practices. *Journal of Educational Assessment, Evaluation and Accountability*, 28: 105-109. DOI: 10.1007/s11092-016-9241-1.
- Ilgén, D.R., Fisher, C.D., & Taylor, M.S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4): 349-371. DOI: 10.1037/0021-9010.64.4.349.
- Ingvarson, L. & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education*, 52(1): 5-35.
- Interstate Teacher Assessment and Support Consortium (InTASC). (2011). InTASC model core teaching standards: A resource for state dialogue. Washington, DC: Council of Chief State School Officers.
- Jawahar, I.M. (2010). The mediating role of appraisal feedback reactions on the relationship between rater feedback-related behaviors and ratee performance. *Group & Organization Management*, 35(4): 494-526. DOI: 10.1177/1059601110378294.
- Johnson, J. (2015). Getting your message out (and why it's not enough). *Educational Leadership*, 72(7): 10-16.
- The Joint Committee on Standards for Educational Evaluation (JCSEE). (2009). *The personnel evaluation standards: How to assess systems for evaluating educators*, 2nd Ed. Thousand Oaks, CA: Corwin Press.

- Josselson, R. (2007). Narrative research and the challenge of accumulating knowledge. *Narrative, state of the art*. Michael G.W. Bamberg, Ed. Benjamins Publishing Co.
- Kampen, J.K. & Tamás, P. (2014). Overly ambitious: Contributions and current status of Q methodology. *Quality and Quantity: International Journal of Methodology*, 48: 3109-3126. DOI: 10.1007/s11135-013-9944-z.
- Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- Kincheloe, J.L. (1993). *Toward a critical politics of teacher thinking: Mapping the postmodern*. Westport, CT: Bergin & Garvey.
- Kincheloe, J. (1997). Fiction formulas: Critical constructivism and the representation of reality. *Representation and the text: Re-framing the narrative voice*. William G. Tierney and Yvonna S. Lincoln, Eds. Albany, NY: State University of New York Press.
- Kincheloe, J.L. (2006). A critical complex epistemology of practice. *Taboo, Fall-Winter*: 85-98.
- Kincheloe, J.L. & McLaren, P. (2000). Rethinking critical theory and qualitative research. In *Handbook of Qualitative Research, 2nd ed.* Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Khachatryan, E. (2015). Feedback on teaching from observations of teaching: What do administrators say and what do teachers think about it? *NASSP Bulletin*, 99(2): 164-188. DOI: 10.1177/0192636515883716.
- Killion, J. (2015). *The feedback process: Transforming feedback for professional learning*. Oxford, OH: Learning Forward.
- Kim, J-H. (2016). *Understanding narrative inquiry*. San Francisco, CA: SAGE Publications, Inc.
- Kimball, S.M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*. 16(4): 241-268.
- Kinicki, A.J., Prussia, G.E., Wu, B.J., & McKee-Ryan, F.M. (2004). A covariance structure analysis of employees' response to performance feedback. *Journal of Applied Psychology*, 89(6): 1057-1069. DOI: 10.1037/0021-9010.89.6.1057.
- Kise, J. (2014). Two sides of the coin: Evaluation and support work together to strengthen teaching. *JSD: The Learning Forward Journal*, 35(6): 24-28.
- Kluger, A.N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2): 254-284.

- Kluger, A.N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, 7(3): 67-72. <http://www.jstor.org/stable/20182507>.
- Kraft, M.A. & Gilmour, A.F. (2017). Revisiting *The Widget Effect*: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5): 234-249. DOI: 10.3102/0013189X17718797.
- Lavigne, A.L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record*, 116:1-29.
- Leech, N.L. & Onwuegbuzie, A.J. (2009). A typology of mixed methods research designs. *Quality and Quantity*, 43: 265-275. DOI: 10.1007/s11135-007-9105-3.
- Le Fevre, D. M. & Robinson, V. M. J. (2015). The interpersonal challenges of instructional leadership: Principals' effectiveness in conversations about performance issues. *Educational Administration Quarterly*, 51(1): 58-95.
- Lincoln, Y.S. (1997). Self, subject, audience, text: Living at the edge, writing in the margins. *Representation and the text: Re-framing the narrative voice*. William G. Tierney and Yvonna S. Lincoln, Eds. Albany, NY: State University of New York Press.
- Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: SAGE.
- Lincoln, Y.S. & Guba, E.G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Lipton, L. & Wellman, B. (2013). *Learning-focused supervision: Developing professional expertise in standards-driven systems*. Charlotte, VT: MiraVia, LLC.
- Lipton, L., Wellman, B., & Humbard, C. (2003). *Mentoring Matters: A practical guide to learning-focused relationships*, 2nd ed. Sherman, CT: MiraVia, LLC.
- Long, H. (2017). Validity in mixed methods research in education: The application of Habermas' critical theory. *International Journal of Research & Method in Education*, 40(2): 201-213. DOI: 10.1080/1743727X.2015.1088518.
- Long, H. & Rodgers, C.R. (2017). Re-conceptualize mixed methods research: Introducing a new conceptual framework. *Quality and Quantity*, 51: 2813-2829. DOI: 10.1007/s11135-016-0447-6.
- Lortie, D.C. (1975). *Schoolteacher: A sociological study*, 2nd Ed. Chicago, IL: The University of Chicago Press.

- Lowenhaupt, R. (2014). The language of leadership: Principal rhetoric in everyday practice. *Journal of Educational Administration*, 52(4): 446-468.
- Luyt, R. (2012). A framework for mixing methods in quantitative measurement development, validation, and revision: A case study. *Journal of Mixed Methods Research*, 6(4): 294-316. DOI: 10.1177/1558689811427912.
- Marchant, G.J., David, K.A., Rodgers, D., & German, R.L. (2015). State teacher evaluation and teacher education. *The Teacher Educator*, 50(2): 89-108. DOI: 10.1080/08878730.2015.1011943.
- Marshall, K. (2013). *Rethinking teacher supervision and evaluation: How to work smart, build collaboration, and close the achievement gap*, 2nd ed. San Francisco, CA: Jossey-Bass.
- Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, November 2012:14-19.
- Marzano, R.J. (2017). *The new art and science of teaching*. Bloomington, IN: Solution Tree Press.
- Marzano, R.J. & Toth, M.D. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. Alexandria, VA: ASCD.
- Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.
- McKeown, B. & Thomas, D.B. (2013). *Q Methodology*. Thousand Oaks, CA: SAGE Publications, INC. online book: DOI: <https://dx.doi.org/10.4135/9781483384412>.
- McLaughlin, M.W. (1990). Embracing contraries: Implementing and sustaining teacher evaluation. In *The New Handbook of Teacher Evaluation*, Jason Millman & Linda Darling-Hammond, Eds. SAGE Publications Ltd.: 403-415. DOI: <http://dx.doi.org/10.4135/9781412986250>.
- MET Project. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation.
- MET Project. (2014). *Building trust in observations: A blueprint for improving systems to support great teaching*. Seattle, WA: Bill & Melinda Gates Foundations. http://collegeready.gatesfoundation.org/wpcontent/uploads/2015/12/MET_Observation_Blueprint.pdf

- MET Project. (2015). *Seeing it clearly: Improving observer training for better feedback and better teaching*. Seattle, WA: Bill & Melinda Gates Foundation.
http://collegeready.gatesfoundation.org/wpcontent/uploads/2015/10/MET_Seeing_It_Clearly_v2.pdf
- Mette, I.M., Anderson, J., Nieuwenhuizen, L., Range, B.G., Hvidston, D.J., & Doty, J. (2017). The wicked problem of the intersection between supervision and evaluation. *International Electronic Journal of Elementary Education*, 9(3): 709-724.
- Merriam, S.B. & Bierema, L.L. (2014). *Adult learning: Linking theory and practice*. San Francisco, CA: Jossey-Bass.
- Merriam, S.B., Caffarella, R.S., & Baumgartner, L.M. (2007). *Learning in adulthood: A comprehensive guide, 3rd Ed.* San Francisco, CA: Jossey-Bass.
- Muñoz, M.A., Scoskie, J.R., & French, D.L. (2013). Investigating the “black box” of effective teaching: the relationship between teachers’ perception and student achievement in a large urban district. *Education Assessment, Evaluation and Accountability*, 25: 205-230. DOI: 10.1007/s11092-013-9167-9.
- Myung, J. & Martinez, K. (2013). Strategies for enhancing the impact of post-observation feedback for teachers. [Brief] Stanford, CA: Carnegie Foundation for the Advancement of Teaching. www.carnegiefoundation.org.
- National Board for Professional Teaching Standards (NBPTS). (2002). What teachers know and should be able to do. Arlington, VA: National Board for Professional Teaching Standards.
- Natriello, G. (1990). Intended and unintended consequences: Purposes and effects of teacher evaluation. In *The New Handbook of Teacher Evaluation*, Jason Millman & Linda Darling-Hammond, Eds. SAGE Publications Ltd.: 36-45. DOI: <http://dx.doi.org/10.4135/9781412986250>.
- Nevo, D. (2006). Evaluation in education. In *The SAGE handbook of evaluation*. Ian F. Shaw, Jennifer C. Greene, & Melvin M. Mark, Eds. SAGE Publications Ltd: 442-460. DOI: <http://dx.doi.org/10.4135/9781848608078.n20>
- Newman, I., Lim, J., & Pineda, F. (2013). Content validity using a mixed methods approach: Its application and development through the use of a table of specifications methodology. *Journal of Mixed Methods Research*, 7(3): 243-260. DOI: 10.1177/1558689813476922.
- New York State Education Department [NYSED]. (2011). *The New York State teaching standards*. University of the State of New York. Retrieved: www.highered.nysed.gov/tcert/resteachers/teachingstandards.html

- New York State Education Department [NYSED]. (2018). Guidance on New York State's annual professional performance review for teachers and principals to implement Education Law §3012-d and the commissioner's regulations. The University of the State of New York: updated April 17, 2018.
- New York State Education Department [NYSED]. (2018). The Directory of Public and Non-Public Schools and Administrators for the State of New York: NYS Administrators Listing, EXCEL-97 SpreadSheet. Retrieved from NYSED Office of Information and Reporting Services. <http://www.nysed.gov/admin/bedsdata.html>. Accessed April 2018.
- New York State Education Law Article 61§3012-d: Annual Teacher and Principal Evaluations (2015). New York State Board of Regents.
- New York State United Teachers [NYSUT]. (2014) NYSUT's Teacher Practice Rubric * 2014 Edition * Aligned with the New York State Teaching Standards. Accessed at: <https://www.nysut.org/~media/files/nysut/resources/2014/september/nysutteacherpractic erubric2014.pdf?la=en>
- New York State United Teachers [NYSUT]. (2019). Changes to annual professional performance review (APPR). Fact Sheet No: 19-09 (Rev 18-12).
- Olsen, M.R., & Craig, C. J. (2001). Opportunities and challenges in the development of teachers' knowledge: The development of narrative authority through knowledge communities. *Teaching and Teacher Education, 17*: 667-684.
- Onwuegbuzie, A.J. & Leech, N.L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology, 8*(5): 375-387. DOI: 10.1080/13645570500402447.
- Onwuegbuzie, A.J. & Johnson, R.B. (2006). The validity issue in mixed research. *Research in the Schools, 13*(1): 48-63.
- Onwuegbuzie, A.J., Bustamante, R.M., & Nelson, J.A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research, 4*(1): 56-78. DOI: 10.1177/1558689809355805.
- Onwuegbuzie, A.J., Johnson, R.B., & Collins, K.M.T. (2011). Assessing legitimation in mixed research: A new framework. *Quality and Quantity, 45*: 1253-1271. DOI: 10.1007/s11135-009-9289-9.
- Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Scotts Valley, CA: CreateSpace Independent Publishing. ISBN-13:978-1500594343.
- Papay, J.P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1): 123-143.

- Patton, M.Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*, 4th Ed. Thousand Oaks, CA: SAGE Publications, Inc.
- Peng, J-C. & Lin, J. (2016). Linking supervisor feedback environment to contextual performances: The mediating effect of leader-member exchange. *Leadership & Organization Development Journal*, 37(6): 802-820.
- Peterson, C. (2017). Exploratory factor analysis and theory generation in psychology. *Review of Philosophy and Psychology*, 8: 519-540. DOI: 10.1007/s13164-016-0325-0.
- Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin*, 88(639): 60-79.
- Phillips, D.C. (1997). Telling the truth about stories. *Teaching and Teacher Education*, 13(1):101-109. DOI: S0742-051X(96)00030-3.
- Polkinghorne, D.E. (2007). Validity issues in narrative research. *Qualitative Inquiry*, 13(4): 471-486. DOI: 10.1177/1077800406297670
- Ponticell, J.A. & Zepeda, S.J. (2004). Confronting well-learned lessons in supervision and evaluation. *National Association of Secondary School Principals (NASSP) Bulletin*, 88(639): 43-59.
- Popham, W. J. (2013). *Evaluating America's teachers: Mission impossible?* Thousand Oakes, CA: Corwin.
- Quintelier, A., Vanhoof, J., & De Maeyer, S. (2018). Understanding the influence of teachers' cognitive and affective responses upon school inspection feedback acceptance. *Educational Assessment, Evaluation and Accountability*, 30: 399-431. DOI: <https://doi.org/10.1007/s11092-018-9286-4>.
- Ramlo, S. (2016). Mixed method lessons learned from 80 years of Q methodology. *Journal of Mixed Methods Research*, 10(1): 28-45. DOI: 10.1177/1558689815610998.
- Range, B.G., Young, S., & Hvidston, D. (2013). Teacher perceptions about observation conferences: What do teachers think about their formative supervision in one US school district? *School Leadership & Management*, 33(1): 61-77. DOI: 10.1080/13632434.2012.724670.
- Reid, D.B. (2017). U.S. principals' interpretation and implementation of teacher evaluation policies. *The Qualitative Report*, 22(5): 1457-1470. Retrieved from <http://nsuworks.nova.edu/tqr/vol22/iss5/19>
- Reilly, M. (2015). Getting genuine commitment for change. *Educational Leadership*, 72(7): 42-46.

- Reio, Jr., T.G. & Shuck, B. (2015). Exploratory factor analysis: Implications for theory, research, and practice. *Advances in Developing Human Resources*, 17(1): 12-25. DOI: 10.1177/1523422314559804.
- Renfro, L. (2014). Turn on the light: Observation tool connects teacher practice with student learning. *JSD: The Learning Forward Journal*, 35(6): 30-34.
- Rigby, J. G. (2015). Principals' sensemaking and enactment of teacher evaluation. *Journal of Educational Administration*, 53(3): 374-392.
- Rosenblatt, P.C. (2003). Interviewing at the border of fact and fiction. *Postmodern Interviewing*. Jaber F. Gubrium and James A. Holstein, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Roussin, J. L. & Zimmerman, D. P. (2014). Inspire learning, not dread: Create a feedback culture that leads to improved practice. *JSD: The Learning Forward Journal*, 35(6): 36-39, 47.
- Saldaña, J. (2013). *The coding manual for qualitative researchers*, 2nd ed. Thousand Oaks, CA: SAGE Publications, Ltd.
- Scheeler, M.C., Ruhl, K.L., & McAfee, J.K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4): 396-407.
- Schmolck, P. (2014). *PQMethod Package: Manual/version 2.35*. available online: <http://schmolck.org/qmethod/>
- Schön, D.A. (1983). *The reflective practitioner: How professionals think in action*. United States: Basic Books, Inc.
- Schön, D.A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.
- Schwandt, T.A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Scriven, M. (1991). *Evaluation thesaurus*, 4th Ed. Newbury Park, CA: SAGE Publications, Inc.
- Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21: 111-129.
- Seidman, I. (2006). Chapter 8-Analyzing, interpreting, and sharing interview material. *Interviewing as qualitative research: A guide for researchers in education and the Social sciences*, 3rd Ed. New York, NY: Teachers College Press.

- Shemmings, D. (2006). 'Quantifying' qualitative data: An illustrative example of the use of Q methodology in psychosocial research. *Qualitative Research in Psychology*, 3: 147-165. DOI: 10.1191/1478088706qp060oa.
- Spry, T. (2001). Performing autoethnography: An embodied methodological praxis. *Qualitative Inquiry*, 7(6): 706-732.
- Steinberg, M.P. & Donaldson, M.L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. Policy Brief: *Association for Education Finance and Policy*: 340-359. DOI: 10.1162/EDFP_a_00186.
- Stephenson, W. (1977). Factors as operant subjectivity. *Operant Subjectivity*, 1, 3-16.
- Stephenson, W. (2014). General theory of communication. *Operant Subjectivity: The International Journal of Q Methodology*, 37(3): 38-56. DOI: 10.15133/j.os.2014.011.
- Stone, D. & Heen, S. (2014). *Thanks for the feedback: The science and art of receiving feedback well*. New York, NY: Penguin Books.
- Strega, S. (2015). The view from the poststructural margins: epistemology and methodology reconsidered. In Susan Strega & Leslie Brown, eds., *Research as resistance: Revisiting critical, indigenous, and anti-oppressive approaches*, 2nd Ed. Toronto, Ontario: Canadian Scholars' Press Inc.
- Stronge, J. H. (2007). *Qualities of effective teachers*, 2nd ed. Alexandria, VA: Association for Supervision and Curriculum Development.
- Stufflebeam, D.L. & Sanders, J.R. (1990). Using the personnel evaluation standards to improve teacher evaluation. In *The New Handbook of Teacher Evaluation*, Jason Millman & Linda Darling-Hammond, Eds. SAGE Publications Ltd.: 416-428. DOI: <http://dx.doi.org/10.4135/9781412986250>.
- Tashakkori, A. & Teddlie, C. (2010). Putting the human back in "Human research methodology": The researcher in mixed methods research. *Journal of Mixed Methods Research*, 4(4): 271-277. DOI: 10.1177/1558689810382532.
- Teddlie, C. & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, 13(1): 12-28.
- Teddlie, C. & Tashakkori, A. (2012). Common "Core" characteristics of mixed methods research: A critical review of critical issues and call for greater convergence. *American Behavioral Scientist* 56(6):774-788. DOI: 10.1177/0002764211433795.
- Tedlock, B. (2000). Ethnography and ethnographic representation. In *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.

- Tuytens, M. & Devos, G. (2012). Importance of system and leadership in performance appraisal. *Personnel Review*, 41(6): 756-776. DOI: <https://doi.org/10.1180/00483481211263692>.
- Tuytens, M. & Devos, G. (2014). How to activate teachers through teacher evaluation? *School Effectiveness and School Improvement*, 25(4): 509-530.
- Tuytens, M. & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers and Teaching: Theory and Practice*, 23(1):6-24. DOI: 10.1080/13540602.2016.1203770.
- Vidich, A.J. & Lyman, S.M. (2000). Qualitative methods: Their history in sociology and anthropology. *Handbook of Qualitative Research*, 2nd ed. Norman K. Denzin and Yvonna S. Lincoln, Eds. Thousand Oaks, CA: SAGE Publications, Inc.
- Wassermann, S. (2015). What is teaching? Inside the black box of what teachers do. *Childhood Education*, 91(2): 83-89. DOI: 10.1080/00094056.2015.1018772.
- Watts, S. & Stenner, P. (2005). Doing Q methodology: Theory, method and interpretation. *Qualitative Research in Psychology*, 2: 67-91. DOI: 10.1191/1478088705qp022oa.
- Weems, D.M. & Rogers, C.B.H. (2010). Are US teachers making the grade? A proposed framework for teacher evaluation and professional growth. *Management in Education*, 24(1): 19-24. DOI: 10.1177/0892020609354959.
- Wiggins, G. (2012). 7 keys to effective feedback. *Educational Leadership*, September:11-16.
- Willis, C., & Ingle, W. K. (2015). "Who has time for this?" Negotiating roles in instructional supervision and evaluation. *Journal of Cases in Educational Leadership*, 18(1): 66-76.
- Wright, P.N. (2013). Is Q for you?: Using Q methodology within geographical and pedagogical research. *Journal of Geography in Higher Education*, 37(2): 152-163. DOI: 10.1080/03098265.2012.729814.
- Zimmerman, S. & Deckert-Pelton, M. (2003). Evaluating the evaluators: Teachers' perceptions of the principal's role in professional evaluation. *National Association of Secondary School Principals. NASSP Bulletin*: 87(636): 28-37.
- Zingoni, M. (2017). Motives in response to negative feedback: A policy-capturing study. *Performance Improvement Quarterly*, 30(3): 179-197. DOI: 10.1002/piq.21249.

Danielle R. Brain

302 Maple Street
Rome, New York 13440
(315) 339-5638
DBrain@twcny.rr.com
drbrain@syr.edu

Professional Profile:

- Professional practice reflects knowledge and application of research-based strategies for student learning and achievement
- Experience with data analysis and collaboration with colleagues for implementing data-driven instructional practices
- Provider/facilitator of professional development with a focus on teacher efficacy and reflective practice

Education:

- Syracuse University: School of Education **December 2020**
PhD: Teaching and Curriculum GPA: 3.93
- Union Graduate College **June 2011**
Certificate of Advanced Study: NBC and Teacher Leadership GPA: 4.00
- State University at Binghamton **May 1990**
Master of Science: Teaching/English GPA: 3.54
- Utica College of Syracuse University **May 1988**
Bachelor of Arts: English/Education GPA: 3.19

Teaching Experience:

- **Utica City School District:** **November 1991 to Present**
Thomas R. Proctor Senior High School:
Advanced Placement English Literature and Composition
English Regents Level 9-11
Academic Intervention Service—ELA Lab 9-12
J.F.K. MS/James H. Donovan MS
General English 7-8
- **Utica College (of Syracuse University)** **Fall 1990 to 2011**
Co-Director, Mohawk Valley Writing Project
Professional Development Coordinator, Recruitment, Continuity
Adjunct Instructor: Division of Continuing Education
Division of Arts and Sciences
English Composition (101, 102, 102w, 201)
Introduction to Literature (135)
Writing Center Evening Supervisor

Certification:

- National Board for Professional Teaching Standards **Initial: November 2001**
National Board-Certified Teacher: **Renewed: October 2010**
Adolescent Young Adulthood/ELA (renewal pending) **October 2021**
- New York State Education Department **February 1995**
Permanent Certification: English 7-12

Organizational Leadership Positions:

- **National Writing Project Site Co-Director:**
 - Responsibilities included NWP grant writing/data collection, establishing-promoting-conducting MVWP professional development programs, building capacity and continuity for MVWP Teaching Fellows
 - Mohawk Valley WP SEED Grant: Site Leadership Summer Institute—2014
 - MVWP/NB Candidate Support Provider: Utica CSD 2012-2013
 - Mohawk Valley Writing Project: Utica College: 2003-2011
 - Empire State Writing Project Network: Leadership Team 2009-2010
 - National Writing Project In-Service Institute: Austin, TX—July 2010
 - National Writing Project Leadership Institute: Oakland, CA-Jan. 2007
- **English Department Chairperson:** Thomas R. Proctor SHS: Utica City School District: 2004-2008
 - Responsibilities included data collection and analysis for ELA Comprehensive Regents examination, developing and maintaining ELA programs and curriculum, facilitating ELA staff/personnel scheduling and accountability
 - Learning Points Curriculum Audit and Co-Interpretation: 2008
 - Title IID Competitive Technology Grant Responsibilities: 2006-2007
- Utica City School District Committees:
 - **Utica Teacher Center Policy Board**
Member: 2013-2018, 2019-present
 - Work closely with UTC Director on program development and building facilitator capacity
 - Attend meetings to approve budgets, grant narratives, and program evaluations
 - **Utica Teachers' Association:**
UTA Representative: 2019-present
Ad Hoc APPR committee: 2016-present
In-Service Credit Hour Approval Committee: 2013-2018
Professional Development Planning Committee: 2013-2017

- Activities as UTA Member focus on forwarding, supporting and monitoring the interests of UTA members on the various committees
- **Other UCSD-Affiliated Activities:**
 - Professional Development Presentations/PLC Facilitator—Spring 2015
 - Utica CSD APPR Teacher Research: Action Research Project—2009-2010
 - Leadership Team: TR Proctor SHS Academy of Business and Finance—2003-2004
 - Utica CSD-New England Small Schools Network Summer Institute—2003
- **New York State English Council (state affiliate for National Council of Teachers of English [NCTE])**
 - Responsibilities included; conducting Business Meetings as Council President; Annual Conference oversight/budgets; newsletter writing and membership updates; aligning activities with NCTE for local national conferences
 - Active Membership on Executive Board as Regional Director, Conference Chairperson, Vice President, and President: 1997-2009
 - Regional Director: New York State English Council, 1997-2009
 - President (Elect-Past): New York State English Council, 2007-2009
 - NYSEC Annual Conference Chair: New York City, November 2007
 - NCTE: Affiliate Leadership Meeting: Chicago, June 2007
 - Welcome Address: National Council for Teachers of English: NYC, 2007
 - NYSED-ELA Summit Meeting: Panel Member from NYSEC, July 2004
 - Vice President (Secondary): New York State English Council, 2000-2003
 - NYSEC Annual Conference Chair: October 2002
 - NCTE Local Affiliate Committee—Registration: Spring Convention, 2000
- **Local/Regional Organizations**
 - Participation in various activities as facilitator for networks or professional development
 - Student Teacher Portfolio Review Committee: SUNY Oneonta, 2005-2012
 - OMH BOCES: Secondary ELA Network Facilitator—2001 to 2010
 - 21st Century Grant: After-school Reading Program: Facilitator: 2006-2008
 - Member: Utica College English Advisory Board, 1997-1999

Professional Development

- **Candidate Support Provider:** National Board for Professional Teaching Standards
 - Provide portfolio support for initial and renewal candidates for National Board Certification across Mohawk Valley region
 - Network with other NYS NBPTS facilitators/community to serve candidates in region
 - Facilitate understanding of certification requirements for candidates
- Related NBPTS programs-professional development:
- MVWP Advanced Institute: National Board Certification CSP, 2010-2011
National Board Candidate Support Provider: Utica CSD—2012-Present
 - NBPTS CSP Training: 2010
- **Doctoral Program:** Syracuse University School of Education
PhD in Teaching and Curriculum: Teacher Evaluation/Professional Development
 - Coursework includes focus on inclusive and urban education
 - Research focus on teacher efficacy, professional development, reflective practice to meet needs of diverse student populations
 - Fulfilling research requirements (IRB approvals) and data collection for program and School of Education projects

Presentations:

- National Writing Project Urban Sites Network Conference: Session Presentation—April, 2015/Culturally Responsive Teaching Practices
- NCTE National Convention: Panel Discussion Leader, November 2011
- Best Practices Conference: Presenter—SUNY Oneonta, October 2010
- New York State United Teachers: Presenter, ELL Conference October 2010
- ELA Summits with NYS Education Department/NYSEC: 2003-2005

Curriculum and Staff Development Activities:

- National Writing Project/Literacy Design Collaborative: Module Writer—2011-2012
- Mohawk Valley WP Advanced Institute: Facilitator Standards Initiative—2010-11
- NYS Education Department: Training/Scoring Facilitator Regents Examinations—1999-2009
- Visual Thinking Strategies: MVWP Program—February 2008
- Reading Comprehension Instruction: AFT ER&D—Summer 2006
- Reading Comprehension Strategies: On-Line Facilitator IBuzz—Spring 2006

- Test Item Review Committee: NYS Education Department, 1999 to 2003
- Test Item Development Committee: NYS Education Department, 2003
- More Effective Schools: Curriculum Development—Summer 1999
- Arts in Education Program for English Language Arts—Summer 1998
- Skillful Teacher Training—Summer 1996
- NYS Humanities Institute: SUNY Albany—Summer 1994

Honors-Awards:

- College Board Advanced Placement Fellowship—2015
- State University of New York at Oneonta: Exemplary Service Award—2010
- New York State English Council: Educator of Excellence—2006