INFLUENCES OF TEACHER DECISION-MAKING AND PROFESSIONAL

LEARNING ON TEACHERS SCORING STUDENT

WRITING ASSESSMENTS

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIRMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF LITERACY AND LEARNING

COLLEGE OF PROFESSIONAL EDUCATION

BY

HEATHER CATO B.S., M.Ed.

DENTON, TEXAS

MAY 2020

DEDICATION

For Papa

Who knew I could do this long before I ever did,

and would have been so proud.

ACKNOWLEDGEMENTS

It is an honor to be express my appreciation for all the people that helped support me in this dissertation journey. Starting a Ph.D. program four days after having giving birth to your first child isn't the most ideal way to kick off an intensive academic journey, so there are many people to thank. First to my husband, who selflessly took care of all the things, including the children and the house while I attended classes, conducted research, and wrote. Simply put none of this would have been possible without your unwavering love and support. To my children, Hudson and Harper, thank you for your snuggles and your words of encouragement. You are both a blessing in my life and have made all of this worth it. Thank you to my parents, grandmother, in-laws and the countless of others who show up at the Thanksgiving dinner table. Each of you have supported me in countless ways every step of this journey and I am eternally grateful for that.

I would also like to thank the supportive professors at Texas Woman's University (TWU), who not only gave their time and expertise in helping me develop as a scholar, but who genuinely took an interest in cultivating meaningful relationships and investing in me on personal level. It is astounding to me to consider the leaps and bounds each professor is willing to take to ensure my academic and professional pursuits are realized. Thank you to my advisor Dr. Pat Watson. Since my undergraduate days at Texas Tech, Dr. Watson has always been my advocate and has diligently pushed me further than I ever thought possible. You were the one that taught me the importance of the balance between research and practice. I can honestly say this dissertation would never have seen

the light of day if you hadn't encouraged me to pursue it. Thank you to Dr. Amy Burke for all of the opportunities, encouragement and guidance. I promise I'll stop preemptively saying my writing isn't any good. Thank you to Dr. Laura Trujillo-Jenks, having someone like you in my corner has been life-saving in more ways than one.

Finally, thank you to the teachers who participated in the study. Your time and willingness to share your voice made this dissertation possible.

ABSTRACT

HEATHER CATO

INFLUENCES OF TEACHER DECISION-MAKING AND PROFESSIONAL
LEARNING ON TEACHERS SCORING STUDENT
WRITING ASSESSMENTS

MAY 2020

In response to the Elementary and Secondary Education Act in 2001 (No Child

Left Behind), states have sought various ways to adequately assess students in order to in

turn hold schools and districts accountable for closing the achievement gap. As the state

of Texas, motivated by pressure from parents, teachers, and other vested interest groups

who expressed concerns over too much testing, explores alternatives to the current

writing assessment system, the purpose of this qualitative study was to explore the

various factors that come into play as a teacher makes scoring decisions while evaluating

student writing. The study considered the teacher knowledge and professional learning

that contribute to the different scoring approaches teachers use while making scoring

decisions. In order to situate the study and consider current efforts underway across the

state, a document analysis was conducted of the current documents related to The Texas

Writing Pilot. The focus of the study was six writing teachers from grades where writing

is assessed across the state (e.g., Grade 4, Grade 7, and Grade 9). Data were collected

through Public Information Requests from TEA, interviews, and think aloud protocols

that captured teachers verbal thinking about his/her scoring decisions while evaluating

student writing.

Findings were presented in three manuscripts written for publication in peer-reviewed journals. These findings revealed a clear disconnect between how educators teach writing and how the state assesses writing. The analysis of the interviews and think-aloud protocol transcripts shed light on the complexity of teacher decision-making. This analysis provided a look into the processes teachers use when making scoring decisions and revealed that teachers do not make scoring decisions in isolation, but rather rely on personal experience, professional learning, and mentorship when making scoring decisions. The findings are a step towards better understanding the influences of teacher decision making when scoring student writing and provide important considerations for a state or educational institution seeking to design assessment with improved inter-rater reliability among educators.

Keywords: writing, writing assessment, teacher decision-making, high-stakes testing, state assessment, student-centered assessment, authentic assessment, alternative assessment, performance assessment, rater agreement, interrater reliability, scorer cognition, rater cognition, analytic writing rubric, accountability, policy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

History often describes educational policy and practice as a pendulum swinging from one extreme idea to the next. While standardized assessments have been part of the educational landscape since the late 20th century, the pendulum of state accountability took a hard swing in 2002 when President Bush reauthorized the Elementary and Secondary Education Act by signing The No Child Left Behind Act (NCLB) of 2001. The goal of this school reform measure required testing of third through eighth-grade students in the areas of literacy and math as a way to hold states and schools accountable for closing the achievement gap (Stiggins, 2002). States had to respond to this new legislation and submit to the federal government new plans for state assessment and accountability. Since that time, states have pursued various ways to adequately assess students to hold schools and districts accountable for their efforts in closing the achievement gap (Darling-Hammond & Adamson, 2010).

In 2013, Texas, like other states seeking to close the achievement gap and meet the demands of No Child Left Behind, reinforced the predominant role tests in play in holding districts accountable for student achievement. In 2013, the 82nd Texas legislative session set into motion requirements for more student accountability, with a new expectation that high school students would have to pass as many as 15 end-of-course exams before being permitted to graduate. By 2015, a mere 2 years later, the 83rd legislative session was forced to respond to the backlash from parents and teachers alike,

who complained about the invasive nature and the sheer volume of state testing. It was this legislative body that began to seek ways to pull back the number of state assessments students were required to take (Montgomery, 2013). As a result, the 83rd Texas Legislature passed legislation, Texas House Bill (H.B.) 866 (2013), which called for a reduction in the number of standardized assessments for students across the state. Motivated to continue to find ways to reduce the number of standardized assessments and increase opportunities for meaningful assessment the 84th Texas Legislature passed Texas HB 1164, which necessitated that the Texas Education Agency (TEA) examine alternative methods for writing assessment by designing and implementing a writing assessment pilot study (Texas Education Agency, n.d.).

Sponsored by Texas House District 1 Representative Gary VanDeaver, and signed into law by then Governor Rick Perry, Texas H.B. 1164 (2015) required the TEA to pilot an alternative state writing assessment (Texas Education Agency, 2016b). Initially, the bill brought before the House removed the written component of the fourth grade, seventh grade, English I and English II writing assessments altogether and put the responsibility for assessing writing back in the hands of local school districts. Revised language of the final bill called for TEA to conduct a study to develop a writing assessment method that would assess:

(1) a student's mastery of the essential knowledge and skills in writing through timed writing samples;

(2) improvement of a student's writing skills from the beginning of the school year to the end of the school year;

(3) a student's ability to follow the writing process from rough draft to final product; and

(4) a student's ability to produce more than one type of writing style (H.B. 1164, 2015).

In response to the legislation, the Student Assessment Division at TEA worked with Educational Testing Service (ETS), the company that holds the state's assessment contract, and representatives from Representative VanDeaver's office to design the parameters for The Texas Writing Pilot. The design of the pilot includes a sampling of students across the state of Texas in Grade 4, Grade 7, English I, and English II completing two timed writing samples (one in the spring and one in the fall) as well as two additional writing samples. The parameters stipulated that these two additional writing samples should mirror classroom writing instruction (Texas Education Agency, n.d.) and incorporate the writing process from start to finish. These samples of student work were collected in portfolios to provide evidence that a student could compose writing in a variety of writing genres. For each student, all four writing samples were given an individual score by their teacher, a blind rater, and a rater at ETS using the same writing rubric. Using a portfolio rubric, each student's portfolio was also given an overall score.

With an intended outcome of the assessments outlined in Van Deaver's bill being an evaluation of a student's growth in writing over the course of a school year, a key feature of Texas Writing Pilot was to provide students with more timely feedback on their writing. The design of The Texas Writing Pilot was to empower the student's classroom

teacher to assess the student writing and provide this timely feedback, as opposed to sending it off for a blind scorer to rate. Ideally, students would then be able to use this specific feedback as they continued to develop their writing skills throughout the year. In order to consider the feasibility of taking such an assessment design to scale across the state, for the purposes of high-stakes testing, TEA conducted a study alongside the pilot to determine score reliability by evaluating, "the quality of locally-produced ratings and whether stakes can be associated with the locally-produced ratings" (Texas Education Agency, 2017, p. 5). Unfortunately, through a feasibility study conducted by TEA, which measured "the quality of locally-produced ratings" (Texas Education Agency, 2017, p. 5), TEA found the results of teachers' scoring was not consistent enough for this type of assessment to be used for the purposes of high-stakes assessment. Taken that the writing pilot was framed around teachers playing a central role in the scoring of student writing, a new challenge was to move beyond scoring reliability, and also consider the various factors that come into play as a teacher makes scoring decisions when evaluating student writing.

Statement of the Problem

It is undeniable that the education system is currently in an age of assessment and accountability (Stecher, 2010). However, in this age of accountability, carefully designed performance assessments can meet federal and state accountability mandates while at the same time providing teachers with valuable instructional data. In order for the states to view assessments such as The Texas Writing Pilot as viable alternative assessments to the current assessment and accountability systems, writing teachers must be able to score

4

student writing using a standardized rubric. From a quantitative research stance, the determination of feasibility for this writing assessment rests squarely with the establishment of inter-rater reliability across raters, yet some researchers suggest that variability is just one factor for consideration in the rating process (Huot, Neill, & Moore, 2010; Jeong, 2015a; Yancey, 1999; Zhang, 2016). For a complete picture, factors such as a teacher's experience and background must also be considered (Jeong, 2015a). Although there is a growing body of research on performance assessments and how raters arrive at their scoring decisions, further study is needed in considering how a teacher's knowledge and training (Birgin & Baki, 2007) contribute to their scoring approach.

Purpose of the Study

As the state explores alternatives to the current writing assessment and determines the scalability of a re-designed assessment option for student writing, this study sought to explore the various factors that come into play as a teacher makes scoring decisions when evaluating student writing. The purpose of this study was to explore the decision-making process when scoring student work samples and to consider the teacher knowledge and professional learning that contribute to the different scoring approaches teachers use.

Research Questions

1. What is the current state of the implementation of H.B. 1164 (2015) regarding performance assessment?

2. What are the various factors that come into play as a teacher makes scoring decisions when evaluating student writing?

a. How do teachers approach the application of a rubric in evaluating student writing?

b. What metacognitive strategies do teachers use in the process of making scoring decisions when evaluating student writing using a rubric?

Significance of the Study

The study contributes to the body of knowledge related to teacher assessment of student writing and adds to the research related to statewide writing assessments. Previous studies have looked at the development and implementation of a statewide writing performance assessment but have only looked at the teacher's ability to evaluate student-writing samples using a rubric in terms of quantitative inter-rater reliability data. While inter-rater reliability is an important feature to consider when determining the scale-ability of teachers evaluating student writing, further inquiry into what influences teachers' scoring decisions is needed. Further, this study contributes to the body of knowledge related to how a teacher's knowledge influences her decision-making. By considering the knowledge and professional learning that contributes to the different approaches, individual teachers use when scoring student writing samples; further light may be shed on features of professional learning that influence teachers' scoring decisions and may subsequently lead to a framework for training teachers as the state looks to scale an alternative form of writing assessment.

Theoretical Framework

We are a collective of our beliefs, experiences, relationships, the way others see us, and the way we see ourselves (Crotty, 1998). As perspective influences the decisions

individuals make (Flick, 2009), in qualitative research, it is essential to frame the study within the researcher's own views of reality, learning, and research. This explanation helped to properly situate the study by providing insight on the types of questions addressed in the study as well as the analysis and discussion of the research.

Knowledge is a constructed endeavor, for it is the individual who possesses the power to construct truth or meaning from their interactions within the world (Crotty, 1998). According to the work of John Piaget, constructivism explains that the construction of knowledge comes through an individual's interactions with the world, people, and things (Piaget, 2014). His work, specifically the "framework for understanding children's ways of doing at different levels of a development" (Ackermann, 2004, p. 3), is often cited in educational arenas. While Piaget's work is often appropriate in educational settings, for this research study, the work of others will be better suited.

After studying with Piaget, Seymour Papert diverted from Piaget's work to highlight the influential role materials, situated within a culture, play in determining meaning-making (Papert, 2006; Papert, 1980; Picard, Papert, Bender, & Blumberg, 2004). While constructionism and constructivism both possess the "connotation of learning as 'building knowledge structures' irrespective of the circumstances of learning" (Harel & Papert, 1991, p. 1), Papert took it one step further. Papert (1980) believed that is that it is impossible for objects to fully be described apart from the person who is experiencing that object because it is the experience the human has with the object that is essential to what the object is and how it is used. As an active participant in the learning

process, new knowledge is developed most effectively by learners when they are "in the process of constructing something external which they can examine for themselves and discuss with others" as they reflect upon the learning and internalize the new knowledge (Picard et al., 2004, p. 262). It is through this hands-on approach that learners have a sense of ownership and personal investment in the learning process.

Whereas Piaget (2014) identified definitive stages of knowledge development in a hierarchical manner, Papert looks at "different approaches to knowledge as styles, each equally valid on its own terms" (Turkle & Papert, 1990, p. 129). Papert's theory of constructionism sheds light on how an individual accesses different media in various contexts in order to form and transform ideas in their mind (Harel & Papert, 1991; Papert, 2006). This idea that an individual can shape and build upon their ideas in context provides a poignant divergence to Piaget's stage theory (Ackermann, 2004; Papert, 2011).

Papert's interest in how a learner engages with various media to boost self-directed learning and support in the development of new knowledge can be applied to teachers as they engage in the act of internalizing professional learning by way of enacting it out in their professional practice. Once teachers experience new learning of any kind, formal or informal, they have an opportunity to return to their classrooms and apply that concept or idea to their professional practice. For teachers, enacting or tinkering with an idea or concept learned, by practicing within the profession, is a way to continue in the act of making knowledge for oneself and improve one's competence in teaching (Duran & Tipping, 2017). For the purpose of this study, the various factors that

come into play when teachers make scoring decisions provided insight into how teachers use professional learning experiences as a mediating tool for internalizing (Vygotsky, 1980) and transforming one's knowledge.

Assumptions

A researcher is able to acknowledge what may be taken for granted relative to the study through the identification of assumptions. For this research, the following assumptions were made. First, teachers of writing also engage in assessing writing and provide feedback to students about their progress. Second, participants will provide honest responses to the survey and interviews.

Definition of Terms

Definitions of the following terms will provide further clarification of meaning within the context of this study.

Accountability-centered assessment. Standardized assessments characterized by a constrained stimulus and structured response (Stecher, 2010) designed to measure student mastery of standards and serve as a data point in state accountability systems to evaluate the performance of educational institutions (Stiggins, 2002).

High-stakes testing. Assessments, standardized or otherwise, are considered high-stakes when consequences, good or bad, are associated with how students perform on the test. Consequences may include individual consequences, such as the student being retained, but could also involve rewards or penalties for the school or the staff of a school (Darling-Hammond, 2008).

Student-centered assessment. Assessment that is centered on the student can be characterized by constructed responses or the production of a product, for the purpose of evaluating student growth and mastery of learning standards or objectives (Stiggins, 2002). In many cases, student-centered assessment is then utilized by an educator to provide descriptive feedback to students as well as make responsive instructional decisions.

Texas Writing Pilot. A writing pilot study conducted by TEA and authorized by Texas H.B. 1164 (2015) was developed to explore an alternative assessment and determine the feasibility of scaling the pilot statewide. The assessments associated with this pilot must take into consideration student mastery of standards and improvement over the course of a school year, a student's ability to follow the writing process, and the ability for a student to write in more than one writing style.

Overview of the Methodology

A qualitative study methodology (Merriam, 2009) was used to explore efforts by the state of Texas to implement an alternative writing assessment as well as to examine closely the approaches individual teachers take when scoring writing. In order to situate the study and consider current efforts underway across the state to determine the teacher's role in assessing alternative writing assessments, a thorough document analysis (Gibson & Brown, 2009) was conducted of the current documents related to The Texas Writing Pilot published by the TEA.

Following the document analysis, purposive sampling (Merriam, 1998) was used to select six Texas writing teachers, two from each of the following grades: Grade 4,

Grade 7, and Grade 9. For each grade level represented, one of the teachers selected

possessed more than five years of experience teaching, and other teacher selected for that

grade level had fewer than five years of experience teaching. Once the six teachers were

selected, each participant received an online participant survey (see Appendix B) to

obtain background information regarding the teacher's experiences and specialized

training in both scoring writing and writing instruction. Each of the participants

scheduled, at a location of their convenience, a face-to-face initial interview (see

Appendix C) that captured the beliefs and opinions of the teacher as they relate to the

scoring and instruction of writing. At the same time as the initial interview, teachers

participated in a think-aloud protocol (see Appendix D) by evaluating student work (see

Appendix G; Appendix H; Appendix I) while thinking aloud about his/her scoring

decisions into an audio recorder. Through this think-aloud protocol (Ericsson & Simon,

1984; Smagorinsky, 1994), effort was made to capture the cognitive process of the

teachers' scoring decisions as well as the teacher's thoughts behind how the decisions are

made (Zhang, 2016). Following the initial interview and think-aloud protocol, the teacher

was given the standardized writing rubric used by The Texas Writing Pilot (see Appendix

F) and asked to score student writing, but this time using the rubric as a guide for

decision-making. The same think-aloud protocol (see Appendix E) was used for

capturing the teacher's cognitive process as scoring decisions are made (Zhang, 2016).

Finally, a follow-up interview (see Appendix J) was conducted to explore the teacher's

decision-making process further while scoring the student writing. Field notes were taken

during interviews, and all interviews and think-aloud recordings were transcribed. Data

analysis drew upon the thematic analysis method (Braun & Clarke, 2006). Findings and

emerging themes were checked with the participants (Lincoln & Guba, 1985).

Organization of the Study

Chapter One provides insight into the background and rationale for the study. It

outlines the purpose and the research questions explored throughout the study. Chapter

Two provides a review of the literature by providing a synthesis of critical topics and

ideas, including assessment and accountability, teacher decision-making, and

professional learning. The methodology of the study, by describing the research design,

instruments utilized to collect the data, as well as the data analysis techniques utilized

will be described in Chapter Three. Chapter Four presents the findings of the study in

three manuscripts written for publication in peer-reviewed journals. Chapter Five

summarizes key findings of the study, explores potential implications, and considers

recommendations for further study References and appendices will follow Chapter Five.

Summary

Even in today's age of assessment and accountability, there is a need for the

educational community to find a balance between assessment for accountability purposes

and assessment to inform instruction. Likewise, as the state looks for alternatives to

traditional writing assessment, including possibly scaling The Texas Writing Pilot, there

is a need to consider the knowledge and training required for teachers to be able to score

writing for the purpose of state assessment and accountability. The purpose of this study

is to consider the state's current efforts for alternative assessment as well as to explore

the scoring decisions teachers make when evaluating student writing and to consider the

various factors at play when these decisions are made. This chapter serves as an

introduction to the purpose, significance, and theoretical framework of the study. In the

following chapters, the concepts introduced in this chapter will be further explored.

CHAPTER II

REVIEW OF LITERATURE

The purpose of this qualitative study (Merriam, 2009) was to explore the various factors that come into play as a teacher makes scoring decisions when evaluating student writing. The context of this study examines the current efforts being made at the state level for teacher evaluated student writing as well as looks closely at six writing teachers, with a diverse range of experiences, who will score student writing samples using a think-aloud protocol. Ultimately, the findings from this study may inform our understanding of the features of professional learning that influence teachers' scoring decisions and subsequently provide a framework for training teachers as the state looks to scale, statewide, an alternative writing assessment.

While exploring the various factors that come into play as a teacher makes scoring decisions when evaluating student writing, it is important to consider how teachers learn or acquire the knowledge and skills on which they base their decision. It is also important to consider the way individuals situate themselves within a greater social and cultural context to participate in the interdependent process of constructing knowledge (Crotty, 1998; Papert, 2011; Piaget, 2014; Vygotsky 1980). An educator, attending a professional learning experience is not a sponge waiting to soak up knowledge from the facilitator. Passive consumption of the information presented is one thing, but to actively construct knowledge, a fundamental consideration must be the social interaction between the learner and others participating in the learning experience

(Bruner, 1990; Cunningham & Duffy, 1996; Vygotsky, 1980). In other words, the way an experience is understood is in direct concert with the manner in which the individual constructs the context of that situation while under the influence of social factors (Crotty, 1998; Flick, 2009; Papert, 2011; Rogoff, 1994; Vygotsky 1980).

This literature review is presented in three sections. The first section will focus on the assessment and accountability of writing by highlighting the history of writing assessments as well as the history of statewide writing performance assessments. The second section of this review will highlight influences on teacher decision-making. The final section of the review will explore ways in which teachers acquire new knowledge by further exploring professional learning for teachers. Each section is presented for the purpose of considering the ways in which teachers use learning experiences as a mediating tool for internalizing (Papert, 2011; Vygotsky, 1980) and transforming one's knowledge to better understand the decisions teachers make when evaluating student writing using a rubric.

Assessment and Accountability

While the current conversation in education is dominated by assessment and accountability, it is not much of a new conversation. Specifically, for the discipline of writing or composition, assessment has always been intertwined with instruction (Elliot, 2005; Huot, 1990; Yancey, 1999). Throughout the evolution of assessment, the prevailing trend has been a call for higher standards, measuring those standards in the form of assessment, and then holding teachers accountable for high levels of student achievement

15

on those assessments (Stiggins, 2002). This section will trace current assessment trends from the 19th century and shed insight into the continual quest for validity and reliability of large-scale assessments (Elliot, 2005; B. Huot et al., 2010; Huot, 1990; Yancey, 1999).

Accountability-Centered Writing Assessment

In the late 19th century, initial efforts regarding assessment for accountability purposes focused on what students needed to do at the university and subsequently, what the secondary schools needed to do to prepare them for university. To move away from the study of Latin and Greek to the study of English composition, Harvard President Charles Eliot instituted a written examination required for admission to the university's composition program (Elliot, 2005; Hamp-Lyons, 2002; Moore, O'Neill, & Huot, 2009). With a focus on correct spelling, punctuation, and grammar, as well as response to a classic work (Elliot, 2005), this style of examination was soon adopted by other universities as a way to "foster educational and structural changes at both the secondary and postsecondary levels" (Moore et al., 2009, p. 111) and ensure students were ready for the rigors of the universities' academic programs. By the turn of the century, the College Entrance Examination Board (CEEB) was established in response to concerns from the preparatory schools that had to prepare students for numerous different exams.

The CEEB's work focused on ways to standardize the assessments across the various universities (Elliot, 2005) by establishing a framework based on the work of Wilson Farrand, President of the Schoolmasters Association of New York and Vicinity, which outlined a rationale and structure for the organization. Of the key points in

Farrand's framework, one guideline from the framework that was not included in the creation of the CEEB was having high school teachers participate in the process of certifying a student's academic achievement. The decision to not include teachers is important because it brings to light that for over 100 years, "a test was assumed to be a better predictor of student success than a teacher's judgment" (Moore et al., 2009, p. 113) for making decisions regarding post-secondary readiness. Moreover, while questions of test reliability began to permeate the conversation of writing assessment even early on in the development of these assessments, it was not until the 1920s that the idea of validity was even considered or discussed as a component of a viable assessment (Diederich, French, & Carlton, 1961; Hamp-Lyons, 2002; Huot, 1990; Huot et al., 2010). A review of reliability and validity as it relates to writing assessment offers further insight into the current testing landscape.

      Reliability. As far as testing is concerned, reliability can be defined as the frequency to which scores from an assessment would be expected to be similar across multiple iterations of the same assessment (Huot et al., 2010; Lemann, 2000; Moss 1994). Whereas instrument reliability refers to a test's ability to produce consistent scores, inter-rater reliability refers to the agreement between raters on the same papers for a given assessment (Huot et al., 2010). One consideration for the widespread adoption of assessments, and more specifically, standardized assessments, can be attributed to a perceived sense of the reliability of the assessment (Moss, 1994).

For the universities, who wanted to ensure they were enrolling the best and the brightest into their schools, and weeding out those who did not belong, these examinations produced results that they counted on for their decision-making. For those involved in the creation of the early assessments, the term reliability became more about the inter-rater reliability (Diederich et al., 1961) than about ensuring the "consistency of a measure to yield the same results in different administrations" (Huot et al., 2010, p. 496). A mathematical formula developed by Charles Spearman in the 1890s allowed test creators to ensure that tests were reliable, or mathematically similar. His work led to research documenting that teachers were unable to have any kind of congruence in grading the same papers (Huot et al., 2010).

Additionally, as the number of students seeking to test rapidly increased, there became a need to find more efficient ways to assess the written text but preserve the integrity of reliability (Hamp-Lyons, 2002). During World War I, the army adopted a standardized assessment known as the Army Alpha test (Huot, 2002). Other assessments, including the SAT, were established as a result of the successful reliability of the army's test and widely adopted because of the efficiency of scoring and strong statistical correlations (Huot et al., 2010; Lemann, 2000). In 1942, the CEEB scrapped the written composition exam altogether and replaced it with a multiple-choice assessment, which, as they saw it, remedied any of their inter-rater reliability issues (Huot, 2002). This effort for objective standardization continued to be the prevailing emphasis on assessment as

tests during the 1970s focused on multiple-choice questions of usage, vocabulary, and grammar (Huot, 2002; Lemann, 2000; Yancey, 1999).

In the latter half of the 20th century, Charles Cooper (1977) examined the effectiveness of scoring, which focused on the overall impact of a piece of writing, to rank students' writing. As a way of validating the scoring process, Cooper identified seven types of holistic evaluation, including general impression marking, formative response, and analytic scoring (Cooper, 1977). Based on his work, Cooper (1977) believed it was possible to improve reliability to acceptable levels when raters shared similar backgrounds and were carefully trained. Training for raters should include not only a focus on rater practices, but also clear explanations for each criterion within the rubric and examples for each descriptor (Jeong, 2015b). Later studies revealed that by making teachers aware of scoring inconsistencies, teachers begin to adjust their scoring, and in turn, their scores become more reliable (Coffman, 1971). "While scoring reliability in writing assessment is undeniably important, it has been equally challenging to deliver" (Huot et al., 2010, p. 500). Although the prevailing trend has been the standardization of assessment, the extent to which tasks, conditions, and scoring are similar for those taking the assessment (Moss, 1994), studies time and time again looking into the effectiveness of standardized assessment, particularly the SAT, have shown that "high school grades are more powerful than any test score" (Sacks, 2000, p. 271).

Validity. Before the 1920s, it was as if validity was something taken for granted (Huot et al., 2010). Because of the simple fact that creators of assessment instruments

were experts of their assessment instruments, it seemed safe to assume that they were also experts on the validity of that assessment as well (Diederich et al., 1961; Huot, 2002). Compounding the lack of clarity regarding the validity, the overemphasis on rater agreement, blurred the difference between validity and reliability (Behizadeh & Engelhard, 2011; Huot et al., 2010; Wiggins, 1993). It was not until 1954 that validity began to be looked at in broader terms than just rater agreement. By 1966 "content, criterion, and construct validity became the three main foci for test validation" (Huot et al., 2010, p. 505).

By the 1980s, some researchers explained the over-reliance on reliability had been because of creating a viable assessment as opposed to defining a theoretical framework for writing assessment (Behizadeh & Engelhard, 2011; Huot, 2002). By working within and against the prevailing psychometric paradigm, researchers such as Edward White decided to confront the issue of validity with writing assessments and set out to "devise a writing test that could meet the standard stipulated by the testing experts" (Yancey, 1999, p. 490). Adapting the then widely accepted testing technology, the newly designed assessment focused on an end of year essay test that was based on the curriculum covered over the course of the year. Three key procedures identified for this assessment, included a writing prompt, anchor papers and scoring guides for raters, and a determination of acceptable agreement by the test-makers, helped to distinguish this assessment from other assessments in the field (Yancey, 1999).

In the 20th century, writing assessment served as a conduit for administrators and politicians who were seeking efficiency and accountability of the educational system (Williamson, 1994). While politicians and other parties invested in the American educational system, still tout efficiency and accountability, a contemporary view of assessment would make the argument that in addition to validity and reliability "test consequences and implications for the local, educational environment [should] be considered as well" (Huot et al., 2010, p. 507). Building upon this survey of writing assessment history, the current political landscape will be further explored.

Current political landscape. Standardized assessments continue to dominate the educational landscape. The current political climate spurred a movement toward measurement-driven instruction (Causey-Bush, 2005). This over-reliance on standardized assessment perpetuates the assumption that these assessments support standards-based reform, will facilitate changes in the school system, and will subsequently improve student achievement (Ellison, 2012). Today, standardized assessments not only measure student aptitude as they did in the early days but achievement on prescribed academic standards as well (Stiggins, 2002). These assessments put pressure on teachers and schools as the results are being used to identify effective and ineffective teachers and schools (Ellison, 2012).

In 2002, with bipartisan support, President Bush reauthorized the Elementary and Secondary Education Act by signing The No Child Left Behind Act (NCLB) of 2001. The goal of this school reform measure, based on the idea of supporting high standards

for all students, was to promote school improvement by requiring testing of third through eighth-grade students in the areas of reading and math as a way to hold states and schools accountable for closing the achievement gap (Darling-Hammond & Adamson, 2010; Stiggins, 2002). Over time, opponents of NCLB believed it had resulted in a reductionist model of education. Teachers often felt they had to sacrifice creativity because of pressure to teach to the test (Darling-Hammond & Adamson, 2010). In writing specifically, the assessments focused on simplistic and basic aspects of the discipline (Dutro, Selland, & Bien, 2013). Typically, assessments created during this time did not focus on higher-order thinking or performance skills but instead focused on multiple-choice items that could be scored quickly and cheaply (Causey-Bush, 2005; Darling-Hammond & Adamson, 2010).

In 2010 as opposition continued to grow with NCLB, and an economic recession set in, President Obama unveiled his Race to the Top initiative that provided competitive grants to states and districts who came up with innovative solutions that would address "urgent improvements in education needed to prepare all students for a globally competitive economy" (United States Department of Education, 2015, p. vi). The four core areas of focus for Race to the Top included:

- the establishment of rigorous academic learning standards aligned to college and career readiness,
- the further development of effective teachers and leaders,

- the creation of data and technology infrastructures to inform and improve instruction, and

- the intentional efforts to turn around low-performing schools (United States Department of Education, 2015).

Positive gains were documented in each of the states that received grants, including improved communication between the state and districts, increased teacher collaboration and support, increased rigorous academic standards, and improved systems to monitor student achievement throughout the course of the year (United States Department of Education, 2015).

To replace the No Child Left Behind Act of 2001, President Obama signed the Every Student Succeeds Act (ESSA), which returned authority to the states by requiring them to establish their own accountability system (Loeb & Hough, 2016). The goals of ESSA promise to continue to make strides towards providing students with a well-rounded education by ensuring all students graduate from high school prepared for college and career, further supporting early education, and reducing standardized assessments that are unnecessary (Obama, 2009).

Student-Centered Writing Assessment

During the late 1980s and early 1990s, "several states and districts across the nation consciously chose to incorporate aspects of higher-order thinking and processing skills into performance-based, alternative assessment systems" (Abbott, 2016, p. 4). A

review of performance assessment, including efforts to implement performance assessment in writing across the nation, provides further context for this study.

Performance assessment. It is one thing to assess what has been taught. It is quite another to assess what has been learned (Wiggins, 1989). Assessment, the measure of one's ability to demonstrate what he or she knows, is not a new concept to education, but rather a recurring hot topic. From when to assess, how often to assess, how to grade, to even the structure of the assessment itself, the perspectives on the matter are quite diverse. In the array of viewpoints on the subject, one type of assessment that continues to be a focus of the conversation is authentic performance assessment (Gomez, 1999; Stiggins, 1987; Tung, 2010; Valdéz Pierce & O'Malley, 1992; Wiggins, 1989).

According to the TEA, "A quality writing assessment is inclusive of multiple factors of development that evidence growth in writing" (Texas Education Agency, 2016c, p. 5). As a type of performance assessment (Stiggins, 1987; Tung, 2010), a writing portfolio diverges from traditional standardized, or multiple-choice, assessment and provides learners feedback on their performance and growth over time in relation to a predetermined continuum of writing proficiency (O'Malley & Pierce, 1996; Valdéz Pierce & O'Malley, 1992). The use of a portfolio assessment for writing has the potential to empower students to systematically collect authentic evidence of work over time that reflects their level of proficiency towards the state's standards for writing (Birgin & Baki, 2007; Gomez, 1999; Stiggins, 1987; Tung, 2010; Valdéz Pierce & O'Malley, 1992; Wiggins, 1989). A key component of performance assessment is often a reflective piece a

student writes to further explain the pieces found in the portfolio (Black & Wiliam, 1998). Additionally, the performance assessment relies on teachers as they utilize rubrics or rating scales to evaluate the students' portfolios (Valdéz Pierce & O'Malley, 1992).

Performance assessments provide students with multiple ways to show what they know (Stiggins, 1987; Tung, 2010). A carefully crafted performance assessment allows learners opportunities for choice as they demonstrate a range of abilities and skills (Tierney, Carter, & Desai, 1991; Tung, 2010). Through this type of assessment, learners construct or create original responses as opposed to the constrained response they are expected to provide on a multiple-choice type of assessment. Typically for performance assessments, evaluation is based on multiple criteria (Gomez, 1999; Wiggins, 1989), often in the form of a rubric, and often includes an actual audience external to the teacher. Performance assessments have the potential to measure cognitive thinking, reasoning, and the ability to problem-solve in meaningful or real-world contexts (Darling-Hammond & Adamson, 2010; Stiggins, 1987; Tung, 2010; Wiggins, 1989). Well-crafted performance assessments are not limited to the recall of memorized information, but provide opportunities for demonstrating growth of essential standards over time as well as the opportunity to collect information about what learners know and are able to do with that knowledge (Wiggins, 1989).

Unlike traditional pencil and paper assessments, performance assessments shed light on student's strengths (Gomez, 1999; O'Neil, 1992; Tung, 2010) and provide teachers with specific information about each student. In her case study of one high

school in Brooklyn, New York, Gomez (1999) found teachers using performance assessments reported that the assessments helped provide ongoing insight into a wide range of a student's skills and abilities. Assessment should be a central component of the learning experience (Wiggins, 1989), for it is through assessment that teachers and learners can engage in dialogue and bring clarity to the learning experience (O'Neil, 1992; Wiggins, 1996). In the classroom, teachers benefit from performance assessments because they can see firsthand a learner's progress and immediately take action to adapt or modify instruction based on what learners demonstrate that they know and can do (Tung, 2010). In the era of NCLB (*No Child Left Behind Act of 2001: Qualifications for Teachers and Professionals*, 2008), state accountability mandates have put pressure on states to expand their assessment efforts. Often states have turned to heavily rely on multiple-choice assessments that can be wholesale distributed and administered relatively inexpensively (Darling-Hammond & Adamson, 2010), but they have not proven to produce the kinds of results for which they were intended.

Statewide portfolio assessments. Fueled by a desire to ensure learners graduate high school equipped to excel in the demands of the global workforce (O'Neil, 1992), many states turned their investigation towards performance assessment. In many cases, a performance assessment in language arts comes in the form of a written constructed response. To encourage changes in curriculum and instruction, both Vermont and Kentucky (Stecher, 1998) have played a central role in the development and implementation of writing portfolios.

Vermont. Driven by educator concern, in 1988, the state of Vermont sought to establish an assessment system that was responsive to the needs of its stakeholders (Valdéz Pierce & O'Malley, 1992). The assessment of writing occurred through an on-demand essay in the fourth and eighth grades that was scored by teachers (Koretz, Stecher, & Deibert, 1992). Through this process, teachers reported that their curriculum was more congruent to state standards (Stecher, 1998). Another benefit of the Vermont model was that teachers were actively involved in creating the assessments and criteria for measuring the assessment. Teachers taking an active role in the construction of assessments proved to be a powerful professional learning opportunity because it was through this experience teachers developed a deeper understanding of what was expected of students (O'Neil, 1992).

Some concerns that arose from the Vermont assessment plan include the nature of the tasks and the feeling that some students were not given sufficient opportunity to demonstrate knowledge of all aspects of the performance task (Koretz, Stecher, Klein, & McCaffrey, 2005). In 1992, 77 principals were interviewed to gauge their experiences with the portfolio program, and although their responses were diverse, the majority of principals brought to light the burdens that their teachers faced implementing the program (Koretz et al., 1992). Specific reasons for the perceived burden included implementation pace, logistics of record-keeping, and the overall demands on time for implementation (Koretz et al., 1992). Ultimately, the burden of developing the portfolios, as well as the

cost of scoring them, proved to be too challenging to continue to sustain at the state level (Tung, 2010).

Kentucky. Unlike the Vermont assessment that had more opportunities for local control, the Kentucky's assessment plan, Kentucky Instructional Results Information System (KIRIS), utilized from 1990–1999, was comprised of a statewide standardized test, a local performance assessment, short performance tasks, and extended time performance tasks that included writing portfolios (Abbott, 2016; Tung, 2010).

For the writing portion of this assessment, the portfolio included six different pieces that collectively were scored holistically (Tung, 2010). A key difference of this assessment model was that training, for many of the state's teachers, took place through an initiative sponsored by the National Writing Project. It is reported that the targeted and intensive professional learning from the National Writing Project significantly contributed to more accurate scoring (Tung, 2010). This professional learning provided teachers the opportunity to engage in the discipline and better understand the theory and research behind the project.

Questions by parents, teachers, and the community alike about KIRIS about the validity of the assessment system as well as philosophical disagreements over the correct way to teach math and literacy ultimately lead to Kentucky adopting a new assessment system, the Commonwealth Accountability Testing System (CATS; Stecher, 2010). While this system retained some of the components of KIRIS assessment, it was a

relatively short-lived assessment system. In the face of NCLB, Kentucky once again

changed its assessment model in favor of a criterion-referenced test (Stecher, 2010).

Student-centered writing assessments, most often in the form of performance

assessments, provide students an outlet to demonstrate what they know in an environment

that supports choice in how a student demonstrates a range of skills and abilities

(Stiggins, 2002). This section served to explore the broad definitions and uses of

performance assessments as well as shed light on work done at a state level to incorporate

performance assessment into the state assessment and accountability system as well as

the ways in which teachers use performance assessments in the classroom. The following

section will explore the way student work within a performance assessment is assessed

through the use of rubrics.

Rubrics

Building on the discussion of performance assessment, this section will explore

the ways in which student work within a performance assessment is scored. In many

cases, performance assessments rely on teachers as they use rubrics or rating scales to

evaluate a students' writing portfolio (Valdéz Pierce & O'Malley, 1992). Rubrics serve as

a qualitative evaluation, based on multiple criteria, (Gomez, 1999; Wiggins, 1989) that

define standards and expectations, as well as important aspects of performance, by

describing the degrees of quality (e.g., from above expectations to below expectations)

for a given task (Jeong, 2015b; Jonsson & Svingby, 2007; Saddler & Andrade, 2014).

Rubrics can be task-specific or generic (Jonsson & Svingby, 2007). In their review of 40

research articles regarding rubrics for writing assessment, Jonsson and Svingby (2007) reported that rubrics were found to enhance scoring reliability, support the sound judgment of performance assessments, and provide opportunities for teachers to promote learning and improve instruction. Typically, rubrics are utilized as either a holistic or an analytic scoring evaluation tool. When scoring holistically, the rater must determine the overall quality of the performance or piece, whereas when scoring analytically, the rater must assign a score to each domain of the rubric (Jonsson & Svingby, 2007). "The effects of a reader's purposes and prior experience raise serious questions about the influences of training on the nature of the reading process" (Huot, 1990, p. 210). While the reason may be attributed to ease and cost efficiency, holistic scoring is usually the preferred method for large-scale assessment (e.g., state and national assessments; Dempsey, PytlikZilling, & Bruning, 2009; Huot, 1990; Lumley, 2002).

In contrast to a holistic rubric, an analytic rubric can provide more specificity and serve as a better diagnostic tool for evaluation within the classroom (Knoch, 2009). In an analytic rubric the important criteria, or features, of the assignment are described in detail, including specificity as to what the assignment looks like, at the high score point, mid score point, and low point score point of the rubric (Cooper, 1977; Dempsey et al., 2009). A score is then given for each domain of the rubric. As an early developer of analytic approaches to writing assessment, Diederich (1974), conducted a study of academic and business professionals as they scored sets of writing papers, on a 9-point scale, without any training or rubric. Diederich (1974) and his colleagues then used the

data collected to tease out recurring themes that they then used to create an analytic framework for writing assessment. The level of specificity provided within the analytic rubric provides an opportunity for teachers to engage students with meaningful feedback as a vehicle for promoting continued learning (Jonsson & Svingby, 2007).

In a 2002 study of raters using an analytic rating scale for scoring ESL students applying to immigrate to Australia, Lumley (2002) wanted to understand how individuals applied features of a rating scale to writing samples. Based on an analysis of a think-aloud protocol conducted by each participant, the data showed that although raters often follow a similar rating process, the scoring decisions made were quite different (Lumley, 2002). While raters attempt to remain close to the language of the scale, the scale is unable to encompass all possibilities within a piece of writing. Although some tension exists with raters between their overall impression of a piece of writing and the specific language of the scale, Lumley (2002) acknowledged it is possible to find consistency in scoring between raters if adequate training and support is provided to the raters.

Scoring using rubrics. Although rubrics are widely used among teachers, the rubrics themselves are often called into question with arguments that they contain inconsistent or vague language within each of the descriptors (Jeong, 2015a). The processes teachers use for rating assessments with rubrics are often criticized as well for being too subjective (Lumley, 2002). As a rater, one must decide, consciously or otherwise, which feature of the scale to pay attention to as well as how to do define and distinguish between the language within the scale of the rubric (Lumley, 2002). Even

when raters make an intentional effort to follow the rubric as written, studies show that invariably the rater's decision-making process can be influenced by the overall impression of the piece of writing, and even personal intuition (Jeong, 2015a). Some studies suggest that rater reliability, in part, is attributed to the training raters receive (Jeong, 2015a; Knoch, Read, & von Randow, 2007). One such study conducted by Jeong (2015b) closely examined experienced teachers' scoring of essays with and without a rubric and compared the data to similar data about novice teachers scoring. Findings from the study indicated that experienced raters were better able to put their personal constructs aside and rely on the rubric for scoring as well as reveal a necessity for rater scoring training that includes detailed descriptor explanations and examples (Jeong, 2015a).

As the validity and reliability of performance assessments continued to be called into question, it appears that rubrics provide a means for supporting validity without sacrificing reliability (Jonsson & Svingby, 2007). And it is through their intentional use that teachers are able to engage in meaningful dialogue with students to promote continued growth as a student.

This section explored current assessment trends, shed light into the role reliability and validity play in large-scale assessments, and took an in-depth look into the way writing assessments are scored using rubrics. The next section will build on performance assessments by considering the decisions teachers make, specifically the decisions they make when scoring writing assessments.

Teacher Decision-Making

At its core, "any teaching act is a result of a decision, either conscious or unconscious" (Shavelson, 1973, p. 144). And while teacher decision-making is a complex balancing act of negotiating one's own beliefs, curricular constraints, and institutional constraints the most foundational skill in teaching is just that – decision-making (Borko, Shavelson, & Stern, 1981; McMillan, 2005; Shavelson, 1973). Studies that focused on instruction have often indicated that experienced teachers are more adept at interpreting, evaluating, and explaining the complexities of classroom instruction, whereas inexperienced teachers rely on more basic or concrete explanations and interpretations of classroom events (McMillan, 2005). While many teacher decisions, especially as they relate to instruction, must be made in the moment, assessment decisions often provide the affordance of time, allowing the teacher to access and consider a wealth of information as part of the decision-making process (Borko et al., 1981).

While investigating classroom assessment and grading decision-making, McMillian and Nash (2000) asked elementary and secondary teachers to explain the assessment and grading procedures they used while scoring student work. In their study, McMillian and Nash (2000) found that often grading assessments involved more individualized principles based on experience and other various influences than common or standards principles of practice. Through analysis, they categorized teacher decision-making into six themes: "teacher beliefs and values, classroom realities, external factors, decision-making rationale, assessment practices, and grading practices" (McMillan, 2005,

p. 35). McMillian and Nash (2000) were then able to identify that the relationship between those categorized could be summarized by the competing forces of internal beliefs with external influences, with teachers' internal beliefs being the most influential of the two for assessment decisions (McMillan, 2005). In this section of the review of the literature, the internal and external factors that influence a teacher's decision-making will be explored.

Internal Factors

Internal factors for decision-making are influenced by an individual's core beliefs and values about children, education, and learning (Gill & Hoffman, 2009; McMillan, 2005; Putnam & Borko, 1997). These beliefs can directly influence our choices, efforts, performance, and even behavior. Beliefs can also develop through one's teaching and learning experiences (Gill & Hoffman, 2009). These beliefs often strongly shape both the instructional and classroom management practices an educator uses in her classroom (Putnam & Borko, 1997), and yet teachers often cannot clearly articulate the beliefs that influence their work (Gill & Hoffman, 2009).

Although teachers' beliefs during the planning process can be hard to measure, by exploring teacher talk during shared planning, Gill and Hoffman (2009), sought to gain insight into teachers' beliefs about teaching and learning by focusing on the rationales behind the teachers' decision-making. Their study consisted of four middle school teachers who were observed once a week, in the fall semester, as they planned, shared lessons, and discussed other pertinent things as a team (Gill & Hoffman, 2009).

34

Ultimately, it was this shared planning that provided insight into the teachers' beliefs about "pedagogical content, general pedagogical beliefs, beliefs about curriculum and textbooks, and beliefs about students' ability" (Gill & Hoffman, 2009, p. 1253).

External Factors

While a teacher's beliefs guide the internal factors of her decision-making, there are a myriad of external forces that may come into play as teachers make decisions. State-mandated assessment, school/district policies, and parents or student home life are just some of the forces that influence teachers as they make decisions on a daily basis in their classroom (Clark & Peterson, 1984; Griffith, Massey, & Atkinson, 2013; McMillan, 2005). For example, a teacher has to weigh the school or district's policies around curriculum, classroom management, or social promotion as she designs lessons, redirects misbehavior, and supports struggling students. A teacher also has to consider the manner in which her students and school will be evaluated. With the external pressures of state-mandated assessment, teachers often feel forced to use more objective assessment practices throughout the course of the school year that mirror the state test in order to ensure their students will be ready for the assessment (McMillan, 2005).

Guided by a widespread assumption that all teachers do now is teach to the test, in her review of studies regarding state-mandated testing, Cimbricz (2003) analyzed the relationship between state assessment and teachers' beliefs and practice. Of the studies reviewed, Cimbricz (2003) found that state testing does, in fact, influence how a teacher behaves and the decisions a teacher makes. However, although state testing does

influence teacher practice and their views of learning, there was not a consistent pattern of influence because other variables such as teacher status or experience level contributed to the ways in which an individual teacher interpreted the importance of the state assessment (2003).

As educators engage with the myriad of decisions to be made during a day, they must carefully weigh these competing forces (Black & Wiliam, 1998; McMillan, 2005). This section sheds light on the internal and external factors that influence teacher decision-making and considers the factors that may come into play when teachers are making scoring decisions.

Professional Learning

In order for teachers to facilitate students to deeper levels of conceptual understanding (Borko, 2004), teachers must have a vast understanding of the discipline for which they teach as well as the research that backs up the practice (Stokes, 2010). Most commonly, opportunities for teachers to develop and deepen their understanding of a given discipline come in the form of professional learning. However, it is no secret that teachers often have a less than enthusiastic outlook when it comes to professional development. In order to ensure that professional learning is both meaningful and relevant, the teachers who participate, the facilitator who guides the learning, and the situation or context for which the professional learning is given must be considered (Borko, 2004).

For a child, learning experiences often focus on the acquisition of skill, but for adult learning, experiences build upon individual's prior knowledge and skill through opportunities that promote self-directed learning (Knowles, 1970, 1996; Merriam, 2001; Merriam, Caffarella, & Baumgartner, 2012; Zmeyov, 1998). In the case of adult learning, learning moves to a more problem-centered approach where teachers are empowered to actively participate in shared inquiry and advocate for their own educational needs (Zmeyov, 1998). With a focus on continuous improvement, the learning is action-oriented and focuses on ensuring high levels of learning for all students (DuFour & DuFour, 2013). Once this new learning is presented or acquired by the teacher, the learning does not stop there. Teachers going back to their classroom to apply this new learning must engage in the act of internalizing professional learning by way of enacting it out in their professional practice. In doing so, teachers draw upon the newly acquired concept or skill and engaged in self-directed learning as a way to develop of new knowledge (Papert, 1980).

In both instances of state-scaled performance assessments previously discussed, Vermont and Kentucky, educator professional development played an important role in sustaining the implementation of the project. In Vermont, the training was intentionally designed to ensure a certain degree of calibration and inter-rater reliability. Over time, it was found that a byproduct of this intensive training was improved instructional practice (Gomez, 1999). Another aspect that contributes to an educator's development of knowledge is the participation in the scoring of the assessments (Darling-Hammond &

Adamson, 2010). Through the collective practice of engaging in discourse and critically scoring the performance assessments within a greater community of teachers, teachers developed a stronger sense of knowledge that was expected for the writing assessment.

Components of Professional Learning

For decades, professional development experts have cautioned against traditional delivery models of professional learning. Teachers need opportunities to engage in just-in-time, job-embedded professional learning (Guskey & Yoon, 2009). When professional development is customized to the needs of teachers and sustained over time, there is evidence that it can positively contribute to improved teacher practice (Borko, 2004). In order for professional learning to prove meaningful for teachers, developers of professional learning must consider conditions for success. Optimal conditions for the development of educator knowledge include: 1) opportunities to study by participating in experiences within the core discipline and reflecting upon that work; 2) collective inquiry with other teachers that focuses on improved classroom practice; and 3) engagement in professional discourse around theory-based research (Darling-Hammond, 2008; Marlink & Wahleithner, 2011; Stokes, 2010). Through meaningful professional learning, teachers can deeply and flexibly develop an understanding of the content so they can help students make connections and address misconceptions (Darling-Hammond, 2008). Teachers are then able to turn their learning around and know how to create experiences in such a way to support meaningful learning among students.

Experience and Reflection

For any given grade and subject area, the complexity and sophistication on the knowledge and skills an educator must possess to teach are immense. College pre-service education preparation programs serve as a beginning for learning the craft of teaching, it is just the beginning, but the adage of practice makes perfect is not just a meaningless phrase. While pre-service programs can provide a solid foundation, they cannot replace the experience one can get by being a teacher. When teachers are able to engage and actually experience (Neuman & Cunningham, 2009) the learning first-hand and then reflect upon it, (Penuel, Fishman, Yamaguchi, & Gallagher, 2007) they are able to have a deeper understanding of the core features and structural components for the given content.

In 2009, Neuman and Cunningham explored the impact of professional development on teacher knowledge for 304 teachers from child-care centers and family daycares in North Carolina. The participants were placed into one of three groups. Group 1 took a three-credit-hour course in early language and literacy; Group 2 also took the course but in addition to the course also received ongoing coaching, and support; and Group 3 served as the control group (Neuman & Cunningham, 2009). Findings from the study indicated it was not merely the professional learning that had the greatest impact on the educator. Instead, it was the reciprocal process of learning and doing that led teachers to be able to implement stronger early literacy instructional practices (Neuman & Cunningham, 2009). When teachers are able to take what they are learning directly back

39

into their classroom and then receive further guidance and support, they are then able to experience the learning firsthand and develop a stronger understanding of how to implement it.

Reflection is also an important factor. Penuel et al. (2007) sampled 454 teachers in an inquiry science program called GLOBE, to determine their ability to take the newly acquired learning and implement it back in the classroom. The authors found that providing teachers time, within the professional learning experience, to create an implementation plan gave the teachers an opportunity to reflect on current district and campus initiatives. The teachers who reflected upon how GLOBE fit within the district and campus initiatives were better prepared to engage students in inquiry strategies they learned from the training.

Collective Inquiry

When teachers engage in a professional learning community (DuFour & DuFour, 2013), collaborating with other teachers and engaging in discourse about student work, they begin to feel empowered to deepen their understanding of the content and ultimately adapt their teaching (Darling-Hammond, 2008). As teachers take part in collective inquiry with other teachers that focuses on improved classroom practice, they are able to develop a shared understanding as they learn from one another (Wilson & Berne, 1999). An important idea related to collective inquiry is that learning is not something to be packaged and delivered to teachers. Instead, teachers should feel empowered to actively

construct their own learning experiences (Wilson & Berne, 1999) as the community

engages in a particular line of inquiry to improve their practice (Borko, 2004).

One key aspect of collective inquiry to consider is the environment and the

culture for which the collective inquiry is given room to grow and develop. While a team

of teachers can easily rally around one another for support, an important player in the

success of collective inquiry would be the support that teachers receive from their

principal. The principal on campus is a pivotal influence on school culture and

subsequently plays an important role in the overall support of the organization's ability to

engage in inquiry. When Barohny and Heining-Boynton (2007) studied influential factors

impacting professional development, they found the school-level support to be important

for teachers with all levels of experience. Specifically, they discovered that there is a

specific need for principals to create supports where teachers can feel empowered to take

advantage of newly acquired instructional methods.

Professional Discourse

Although quite often, teachers seem to prefer to skirt around the conversation of

theory-based research, it is widely known that teachers enjoy talking about issues and

subjects relevant to their work and their students (Wilson & Berne, 1999). Some research

suggests that this professional discourse is something teachers often do not have an

opportunity to develop (Wilson & Berne, 1999) and may emerge as teachers further

engage in collaborative professional communities.

In an effort to increase teacher efficacy, Ross and Bruce created a professional program for 106 sixth grade teachers in Canada. The teachers were randomly assigned to a treatment or a control group (2007). The researchers found that "the contributions to teacher self-assessments, in concert with information on innovative instruction, heighten teacher efficacy, which influences teacher goal setting and effort expenditure" (Ross & Bruce, 2007, p. 52). In this study, a critical component of the professional learning model was an opportunity for teachers to come back together after they have taught a lesson and debrief the experience and share evidence of student learning. According to their findings, teacher efficacy improved when they had an opportunity to engage in a conversation wherein they:

a) saw themselves as an expert or master teacher,

b) saw other teachers being successful,

c) supported each other in believing they could implement the new curriculum, and

d) participated in stress-reduction experiences (Ross & Bruce, 2007).

For Ross and Bruce, while these four components of conversation improved teacher efficacy, it also provided a mediating tool to explain why the strategies utilized in their study "influenced the teachers' beliefs about their effectiveness" (2007, p. 54). When professional learning is inclusive of opportunities to participate in experiences within the core discipline, collective inquiry, and professional discourse around theory-based research, teachers are equipped to develop as professionals.

When teachers have opportunities to participate in experiences within the core discipline, take part in collective inquiry to improve classroom practice, and engage in professional discourse around theory-based research, the conditions are ripe for them to learn and grow as a professional. Through the exploration of these three ideas, this section provides insight into the ways teachers develop their knowledge as an educator, which ultimately is one potential influence of decision-making when scoring writing assessments.

Summary

This review of literature provides an insight into how the history of writing instruction and assessment has been intertwined from the beginning as teachers and lawmakers alike search for ways to have students demonstrate what they know as well as hold schools and teachers accountable. The review provides insight into the ways in which other states have made attempts to adopt a more student-centric type of writing assessment, which may prove valuable as Texas looks to scale an alternative form of writing assessment.

The review of teacher decision-making provided insight into the various decisions that teachers make on a daily basis and the conflicting forces that they consider while making decisions. This information about teacher decision-making further contributes to knowledge about professional learning that was also shared. The review of the literature on professional learning provided insight into the ways in which teachers engage in learning. Because teachers play a pivotal role in student assessment, and the parameters

43

of The Texas Writing Pilot included having teachers assess student writing, insight into

teacher decision-making as well as the ways in which teachers expand their knowledge

through experiences in professional learning was shared.

CHAPTER III

METHODOLOGY

In order to further conversation regarding alternative assessment options for student writing assessment across the state of Texas, within the context of high-stakes testing and accountability, this research study focused on the approaches individual teachers take when scoring writing. A qualitative study methodology (Merriam, 2009), which is situated within a constructivist research paradigm (Guba & Lincoln, 1994), was used to explore the different approaches individuals take when scoring student writing. Through the interactions between individuals and their social worlds, constructivists believe that knowledge is socially constructed (Guba & Lincoln, 1994). The theoretical orientation of this study situates the experience of writing and the evaluation of writing as a social-cognitive process, which cannot be studied outside of its social context (Merriam, 2009).

In considering the knowledge and training that contribute to different approaches a rater takes when scoring, this study examined current efforts by the state of Texas to determine the teacher's role in assessing alternative writing assessments and then explored the knowledge, beliefs, and cognitive strategies of teachers through the examination of multiple sources of evidence (Wolfe & Feltovich, 1994). The following research questions guided the study:

1. What is the current state of the implementation of H.B. 1164 (2015) regarding performance assessment?

2. What are the various factors that come into play as a teacher makes scoring decisions when evaluating student writing?

   a. How do teachers approach the application of a rubric in evaluating student writing?

   b. What metacognitive strategies do teachers use in the process of making scoring decisions when evaluating student writing using a rubric?

In order to explore these research questions, I used a qualitative study methodology (Merriam, 2009). This chapter begins by discussing my research design. Next, I explain the setting and participants, followed by explaining data collection and analysis. Finally, I explain the methods followed to ensure the trustworthiness (Merriam, 2002) of the study.

Research Design

In qualitative research design the researcher, through an inductive process, is the primary instrument to explore meaning making and understanding of their participants or the phenomena being studied (Creswell, 2013; Crotty, 1998; Erlandson, Harris, Skipper, & Allen, 1993). In an effort to explore the various approaches used when scoring student writing, a qualitative study design was selected (Merriam, 2009) as it provides an opportunity to probe into the various ways meaning is constructed. Guided by the research questions (Stake, 1995), the examination of diverse perspectives will provide an

increased depth of understanding of the central phenomenon to scoring student writing and will allow for the collection of various sources of data such as questionnaires, interviews, and observational data (Stake, 1995). The nature of the study and inclusion of data collected from the state and from multiple teachers will support the identification of different approaches individuals use when scoring student writing through an analysis of themes. Guided by the research questions (Stake, 1995), the examination of diverse perspectives provided an increased depth of understanding of the central phenomenon of scoring student writing and allowed for the collection of various sources of data such as surveys, interviews, and observational data (Stake, 1995). The nature of the study and inclusion of data collected from the state as well as from multiple teachers support the identification of different approaches individuals use when scoring student writing through an analysis of themes.

This research study was implemented in four phases as depicted in Figure 1 below. Phase I included a document analysis (Gibson & Brown, 2009) of found artifacts related to The Texas Writing Pilot. Phase II included enrollment of the participants through purposeful sampling (Merriam, 1998) and an initial open-ended participant survey (see Appendix B). Phase III involved data collection through interviews and interaction with the participants selected. Semi-structured, face-to-face interviews with participants, a think-aloud protocol utilized without a rubric, and then with a rubric as participants score writing samples, and a follow-up semi-structured interview was conducted (Merriam, 1998). Phase IV required analysis of participant responses using

47

coding, validation of transcribed information through member checking, and the

identification of emerging themes (Creswell, 2013; Stake, 1995).

Phase I
- Document Analysis of Texas Writing Pilot

Phase II
- Enrollment of participants
- Online Participant Questionaire to obtain background information

Phase III
- Inital face-to-face interview
- Inital think-aloud without using a rubric
- Second think-aloud using a standardized rubric
- Followup face-to-face interview

Phase IV
- Analysis of participants responses
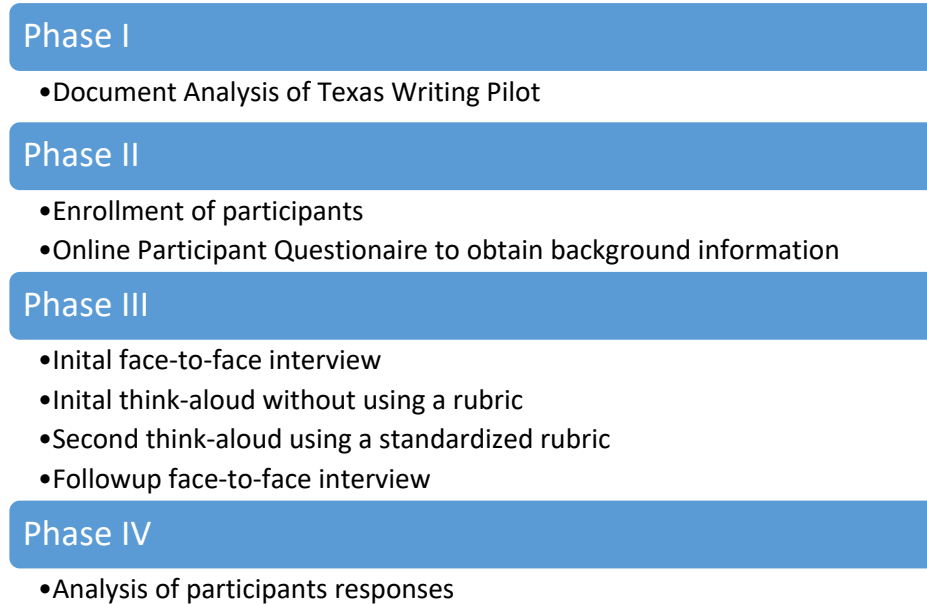
Figure 1. Phases of the Research Process.

Researcher's Role

As a certified Texas public school educator for over 13 years, I have a deep

interest in the ways in which we assess students. While in my experience, I have had the

privilege to teach middle school students, I have also had the opportunity to coach and

lead pre-kindergarten through 12th grade teachers in the field of literacy. It was in my

role as a district administrator for Language and Literacy that I became involved with

The Texas Writing Pilot. Initially, I was working with several other educational leaders

from around the state to develop a portfolio assessment plan that districts could utilize locally. When Texas H.B. 1164 (2015) was signed into law, we turned our efforts to assist the state in developing what would become known as The Texas Writing Pilot. Initially, I played a role in assisting the state with the design specifications for The Texas Writing Pilot as a member of a committee who gave feedback regarding the pilot. During the second year of implementation of the pilot, I began working at one of the regional service centers as the primary liaison for the pilot between the TEA and the participating school districts. In this role, I participated in weekly meetings with members of TEA, the other participating service centers, and the ETS. ETS was the vendor responsible for the development and distribution of state assessments. In this role, I assisted in the implementation of the pilot, sought feedback from my participating districts, provided scoring training to teachers participating in the pilot, and was influential in the redesign of the rubric.

As for my role as the researcher within this study, my role was to observe, interview, collect, and analyze data. Because of my involvement with The Texas Writing Pilot, my study necessitated bracketing throughout the study. Bracketing requires a researcher to bracket or set aside their personal experience, to the greatest extent possible, in order to gain a fresh perspective of the phenomenon being studied (Creswell, 2013). Specifically, when conducting the semi-structured interviews and think-alouds with the participants, I refrained from adding comments or asking questions that were not within the scope of the protocols. While having familiarity of content and context for the study

helped validate my research, bracketing allowed me to gain new insight into the questions being asked in the study.

Setting and Participants

This study took place in the state of Texas in order to explore the efforts being done by the state in the area of alternative writing assessment. Second in total land area only to Alaska, Texas served just short of 5.5 million students during the 2017–2018 school year. From 2007 to 2017, a 10 year period, student enrolment in Texas public schools increased by 15.6%. Of the 5,399,682 students enrolled in Texas schools for the 2017–2018 school year, 52.4% identified as Hispanic, 27.9% as White, 12.6% as African American, 4.4% as Asian, and 2.3% as multiracial. Additionally, the percentage of students identified as economically disadvantage rose from 55.2% in 2007–2008 to 58.7% in 2017–2018. To serve those students in 2017–2018, more than 8,900 schools in 1,200 school districts and open-enrollment charter schools employ more than 356,000 teachers (Division of Research and Analysis Office of Academics, 2018).

Participants for this study included Texas writing teachers from the North Texas area near or around Dallas Fort Worth who have experience teaching writing at a grade level where writing is assessed by the state (i.e., Grade 4, Grade 7, and Grade 9). Purposeful sampling (Merriam, 1998) was used to select six participants selected for this study. Three teachers, one from each of the three grade levels (Grade 4, Grade 7, and Grade 9), had more than 5 years of experience teaching and three teachers, one from each of the three grade levels (Grade 4, Grade 7, and Grade 9) had fewer than 5 years of

experience teaching. Table 1 shows the participants, grade levels taught, and years of experience teaching.

Table 1

Experience of Study Participants

Participants	Grade Level Taught	Years of Experience
Participant 1	9th Grade	More than 5 years
Participant 2	9th Grade	Less than 5 years
Participant 3	7th Grade	More than 5 years
Participant 4	7th Grade	Less than 5 years
Participant 5	4th Grade	More than 5 years
Participant 6	4th Grade	Less than 5 years

This range of experience among the teachers provided insight into the influence these teachers draw upon when making scoring decisions. In addition to the range of years of experience, each of the teachers had a diverse background of experiences and professional learning experiences that contributed to their process for scoring student writing.

In order to find six teachers to participate in this study, I emailed members of MetroCREST, an organization in the North Texas area for district leaders of English Language Arts. I have been a member of MetroCREST for 6 years and have a strong working relationship with many of the members. I knew approaching MetroCREST would be an expedient way to go about gathering potential participants since the

members work closely in their school districts with writing teachers. In the email, I provided information about the study then asked that they forward a survey to experienced and inexperienced writing teachers in the fourth, seventh, and ninth grade who might be willing to participate in my study (see Appendix A).

After the original email was sent, I did not receive any prospective participants, so I followed up by reaching out directly to several members of MetroCREST, whom I knew personally because of my own affiliation as a member of the organization. From those contacts, I was able to secure an experienced and inexperienced writing teacher in each of the fourth, seventh, and ninth grades. IRB approval was secured, and all six participants agreed to participate.

Data Collection

Data Sources

Several data sources were collected for this study to provide rich, thick descriptions for each research question (Creswell, 2013). In order to situate the study and consider current efforts underway across the state to determine the teacher's role in assessing alternative writing assessments, a thorough document analysis (Gibson & Brown, 2009) was conducted of the current documents related to The Texas Writing Pilot published by the TEA. This first phase of the research study included a review of archival data related to The Texas Writing Pilot, which included the Texas Agency's initial request for participants and the interim reports presented to the Texas Legislature by the TEA.

Data collection also involved the use of multiple research methods, including surveys, interviews, observation, and review of collected artifacts (Creswell, 2013) from teachers participating in the study. Multiple data sources were used during data collection to provide a focused yet comprehensive view of the various factors that come into play as a teacher makes scoring decisions when evaluating student writing samples. The primary unit of analysis was teachers' verbal and written statements that reveal their processes of pedagogical reasoning and action (Shulman, 1987) by investigating the different approaches individuals took to score student writing.

To capture and describe these approaches, Figure 2 shows the Process for Data Collection. A pre-scoring participant survey, initial interview, two think-aloud protocols (Ericsson & Simon, 1984; Smagorinsky, 1994), one with and one without a standardized rubric was utilized. Recorded field notes from observations of participants' behavior as well as interactions during think-aloud protocols were recorded to provide additional insight into what the teachers are considering when they score. To triangulate the data (Marshall & Rossman, 2014), post-scoring interviews were conducted, and reflections from the participants were gathered.

| Online Participant Questionaire to obtain background information |
| Inital face-to-face interview |
| Inital think-aloud |
| Second think-aloud using a standardized rubric |
| Followup face-to-face interview |
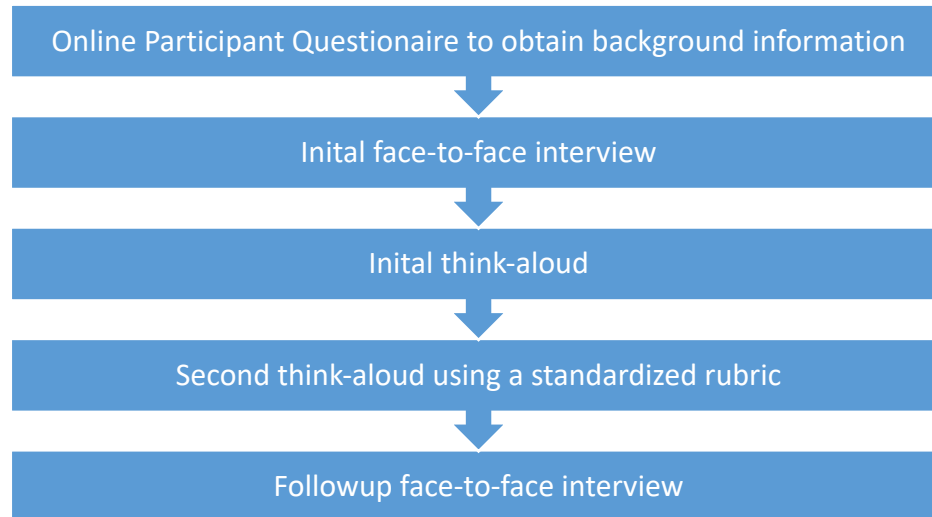
Figure 2. Data collection process.

In order to better understand and consider the sources of knowledge teachers draw from when evaluating student writing I began by asking each participant to complete an open-ended participant survey (see Appendix B). This survey collected initial information about each participant including years of experience, teaching certifications and initial background about their experiences with specialized training in both the scoring and the instruction of writing. Emerging patterns of teacher knowledge and training were explored through the analysis of the teachers' open-ended responses.

Building from this survey, I arranged a time to meet with each teacher face-to-face and engage with them in a semi-structured interview (Merriam, 1998). This interview allowed me to get a better picture of the teachers' beliefs and opinions as they relate to the scoring and instruction of writing (see Appendix C). This interview also

allowed me to probe into not only the teachers' beliefs about experiences in teaching writing as well as their experiences with professional learning related to writing instruction and assessment. In general, the 8-question interview protocol lasted 30 – 40 minutes. The use of the semi-structured interview (Merriam, 1998) protocol supported the consistency of the interviews to ensure sufficient information was collected (Marshall & Rossman, 2014).

During the time of the initial interview, I asked teachers to rate several student writing samples using a think-aloud protocol (Ericsson & Simon, 1984; Smagorinsky, 1994; see Appendix D). These writing samples were original samples from the Texas Writing Pilot training materials (see Appendix G, Appendix H, and Appendix I). The original sample set contained 12 writing samples for each grade level. For the purposes of this research, 6 of the 12 samples were selected. In selecting from the 12 samples, I made sure that a variety of writing, both hand written and typed, were used.

For the first round of scoring student writing, while I did not provide them with a rubric, I did ask that they vocalize their thoughts, explain what they notice about the writing, and provide an explanation for how they might score the essay if they were to give it a grade. My aim was to gain insight about what the participants thought about and considered when making their scoring decision without being encumbered or guided by a rubric. Recordings of the think-aloud were transcribed and coded.

Following the teachers scoring of the student writing without a rubric, I then gave the participants the scoring rubric that was used for The Texas Writing Pilot (see

55

Appendix F). Because all six of the participants had not previously participated in The

Texas Writing Pilot, they had not seen this rubric. I asked the participants to take some

time to get familiar with the rubric, before giving them several more writing samples to

score – this time using The Texas Writing Pilot rubric. As I observed the participants and

heard their comments during the think-alouds, I began to see patterns emerge for the

knowledge teachers draw from when making scoring decisions. Capturing those scoring

decisions and how they were made were used for the purposes of considering the various

factors that come into play as a teacher makes scoring decisions when evaluating student

writing. Recordings of the second think-aloud were also transcribed and coded.

When scoring with a rubric, participants used the scoring rubric of The Texas

Writing Pilot (Texas Education Agency, 2016c) that was collaboratively developed by

teachers, employees at the regional educational service centers, in collaboration with

institutes of higher learning. This rubric consists of four scoring domains: organization,

content, language, and conventions. With the exception of conventions, each of the other

three domains consists of a six-point scale that ranges from Very Limited (1), Limited

(2), Basic (3), Satisfactory (4), Accomplished (5), Exceptional (6). Within each domain,

an explanation of each score point further explains the criteria and expectations as each

level.

After some initial analysis, I met with each participant again for a post-scoring

interview. A transcription of the interviews and the think-aloud protocols were typed and

validated through member checking (Creswell, 2013). In this final interview, I asked each

participant to review the transcription of them scoring the student writing samples. As we reviewed the transcripts together, I would stop to ask them about where they learned a particular idea or concept that they mentioned while scoring. For example, one teacher while scoring remarked about the quality of a students' use of subordinate clauses. When asked how she learned about subordinate clauses she shared about a specific time when she attended a training by Jeff Anderson.

Data Analysis

The purpose of this study was to explore the various factors that come into play as a teacher makes scoring decisions when evaluating student. Data analysis drew upon the thematic analysis method (Braun & Clarke, 2006) to support attainment of the study's purpose. Drawing from the sociocultural perspective (Crotty, 1998; Papert, 2011; Vygotsky, 1980), the analysis was framed by the teachers' participatory learning experiences taking into account both the social and cultural perspectives. For this reason, I paid particular attention to how the views of self or others shaped the participants' decisions. In order to identify themes of the data collected, the data analysis process included the use of coding to sort, organize, and synthesize the collected data to identify emerging themes (Creswell, 2013). Categories were derived from the research questions, the purpose of the study, and data collected and analyzed. The use of coding (Creswell, 2013) provided guidance in understanding and making meaning of the data collected as a result of a thorough analysis of information (Erlandson et al., 1993). Further analysis included a review of patterns found within the codes to identify any additional sub-

categories or overarching themes. Each article presented in Chapter 4 presents further

details the different methods used for data analysis.

Trustworthiness

Bracketing requires a researcher to put their experiences aside, to the greatest

extent possible, in order to take a fresh look at what is being studied so reliability for the

study can be maintained (Crotty, 1998). It should be noted that I did not supervise any of

the teachers who will be participating in the study. However, as an initial contributor to

the design of the state writing pilot, and as an active participant with The Texas Writing

Pilot, I acknowledge the need to bracket my experiences as they relate to the

implementation of the statewide pilot and the use of the scoring rubric. I assert that my

connection and involvement with the statewide pilot study lends credibility and

trustworthiness to the study (Marshall & Rossman, 2014).

Credibility and trustworthiness (Merriam, 2002) throughout the data collection

and analysis process was established and obtained through use of frequent debriefing

sessions between the researcher and peer debriefers, including a district administrator

who has taken part in the Texas Writing Project since inception and, another doctoral

student with expertise in writing instruction and assessment. Additionally, member

checks were conducted through follow-up questioning in the post-scoring interviews.

Summary

The purpose of this qualitative study (Merriam, 2009) was to explore the different

approaches individuals take when scoring student writing. The four phases of 1)

document analysis (Gibson & Brown, 2009), 2) enrollment, 3) data collection, and 4)

data analysis were utilized to explore the knowledge, beliefs, and cognitive strategies of

teachers through the examination of multiple sources of evidence. The selection of

participants and the design of research instruments are aligned with the research

questions and purpose of the study.

CHAPTER IV

PRESENTATIONS OF FINDINGS

The purpose of this study was to explore the decision-making process when scoring student work samples and to consider the teacher knowledge and professional learning that contribute to the different scoring approaches teachers use. An analysis of documents related to The Texas Writing Pilot, the think-aloud protocols conducted by the teachers, and the semi-structured interviews of the teachers provided further insight into influences of teacher scoring as well as how this information contributes to statewide assessment initiatives.

This chapter details the findings revealed from an analysis of the data sources by providing three stand-alone journal articles. The first article written for *English in Texas,* presents the analysis and findings of the document analysis of the found artifacts related to The Texas Writing Pilot. This article for *English in Texas* was written with the expressed intent of answering the first research question: What is the current state of the implementation of H.B. 1164 (2015) regarding performance assessment?

The second journal article presents the data collected from the think-aloud protocols conducted by the teachers as they scored student writing samples without a rubric and then with a rubric. The second article, intended for publication in *Assessing Writing*, sought to consider the research questions:

1. What are the various factors that come into play as a teacher makes scoring decisions when evaluating student writing?

a. How do teachers approach the application of a rubric in evaluating student writing?

b. What metacognitive strategies do teachers use in the process of making scoring decisions when evaluating student writing using a rubric?

Finally, the third article, submitted to *JoLLE@UGA* for publication, synthesized the data collected from the open-ended participant survey, the initial semi-structured, face-to-face interview, and the follow-up semi-structured face-to-face interview in order to dive further into the research question – What sources of knowledge do teachers draw on when evaluating student writing?

Lessons Learned From the Texas Writing Pilot, English in Texas

As an English teacher, and one who particularly loves to teach writing, I have always had some level of internal conflict between how I think I should go about teaching the state standards (in a meaningful way that will help the students see themselves as readers and writers) and how I know students will ultimately be assessed on the state's standardized assessment. Stripping writing down to a single expository essay of 26 lines has always seemed to me to be a limited scope through which students demonstrate their writing skills. So, in 2016, when I was given the opportunity to attend a meeting to talk about the what-ifs and the possibilities of authentic writing assessment, I jumped at the chance. Little did I know, 3 years later, this meeting would ultimately lead to my involvement at the state level to explore the possibilities of scaling portfolio writing assessment across the state.

Initially, I played a role in assisting the state with the design specifications for The Texas Writing Pilot as a member of a committee who gave feedback regarding the pilot. During the second year of implementation of the pilot, I began working at one of the regional service centers as the primary liaison for the pilot between the TEA and participating school districts. In this role, I participated in weekly meetings with members of TEA, the other participating service centers, and the ETS. ETS was the vendor responsible for the development and distribution of state assessments. In this role, I assisted in the implementation of the pilot, sought feedback from my participating districts, and was influential in the redesign of the rubric in year two.

After 3 years of work on the pilot, the state abruptly ended the project. As I reflected upon my involvement with and the results of The Texas Writing Pilot, I still had a number of questions. Why did the state report that the pilot was not a valid assessment instrument? What makes an assessment a strong and valid assessment? How might instruction play a more pivotal role within assessment? How might we, as a community of literacy educators, advocate for authentic and meaningful assessment across the state? But before I explored any of these topics, I wanted to start with simply - what happened?

So, how did we get here? In 2015, there continued to be an outcry from educators and parents across the state to reduce the burden of standardized assessments upon our students. In response, Representative Gary VanDeaver introduced Texas H.B. 1164 (2015), which was later signed into law. The language of the final bill called for TEA to conduct a study to develop a writing assessment method that would assess:

1. a student's mastery of the essential knowledge and skills in writing through timed writing samples;

2. improvement of a student's writing skills from the beginning of the school year to the end of the school year;

3. a student's ability to follow the writing process from rough draft to final product; and

4. a student's ability to produce more than one type of writing style (H. 1164, 2015).

Based on accounts from those who worked on the design of the pilot we know that in response to the legislation, the Student Assessment Division at TEA worked with ETS, the company that holds the state's assessment contract, and representatives from Representative VanDeaver's office to design the parameters for The Texas Writing Pilot. In its final form, the design of the pilot included students in Grade 4, Grade 7, English I, and English II completing two timed writing samples (one in spring and one in fall) as well as two additional writing samples. The timeline for submitting the four writing samples was spread throughout the course the school year. Figure 3 depicts the TEA Timeline for The Texas Writing Pilot.

September	October–December	January–March	April–May	June–July	August
• Initial trainings: Writing Pilot Rubric, online tool • Timed Sample 1	• Ongoing training • Process Sample 1	• Ongoing training • Process Sample 2	• Timed Sample 2 • Local portfolio scoring	• Blind Scoring • Year 3 Planning	• Year 3 Implementation
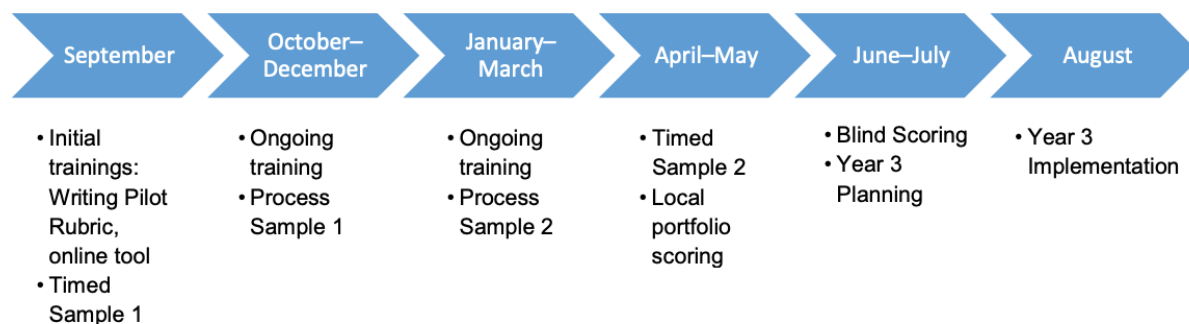
Figure 3. TEA Timeline for The Texas Writing Pilot.

A key feature of Texas Writing Pilot was to provide students with more timely feedback on their writing so they could use that feedback to continue to improve as writers throughout the course of the school year. Because of this, much debate took place over the role the teacher would play in this pilot. Ultimately, to provide timely feedback to the students, the design of the pilot was to empower the student's classroom teacher to assess the student writing as opposed to sending it off for a blind scorer to rate. Ideally, students would then be able to use this specific feedback as they continued to develop their writing skills throughout the year.

The parameters stipulated that these additional writing samples should mirror classroom writing instruction (Texas Education Agency, n.d.) and incorporate the writing process from start to finish. Samples of student work were collected in portfolios to provide evidence that a student could compose writing in a variety of writing genres. For each student, all four writing samples were given an individual score by their teacher and

a blind rater. A sampling of writing samples would be then sent to ETS to receive a third score. Using a portfolio rubric, each student's portfolio was also given an overall score.

In addition to designing and executing the pilot, TEA also conducted a study alongside the pilot to determine score reliability by evaluating, "the quality of locally-produced ratings and whether stakes can be associated with the locally-produced ratings" (Texas Education Agency, 2017, p. 5). The goal was to determine the feasibility of taking such an assessment design to scale across the state, for the purposes of high-stakes testing.

Informally, many administrators and educators praised the pilot. They reported students were writing more and the quality of writing instruction was better because it was unencumbered by the strain of the STAAR test. However, in the final report TEA sent to the governor and the state legislature, TEA, reflecting on both years of the pilot, concluded that the "pilot did not prove to be a valid assessment instrument" (Texas Education Agency, 2018, p. 22). Which again, left me perplexed, wondering how something with so much promise did not pan out as a viable assessment option, at least according to the state.

Methods

Context

The purpose of this study was to review official documents related to The Texas Writing Pilot and explore the state's implementation of H.B. 1164 (2015) as it pertains to student performance assessments. Data were collected using the Texas Education

Agency's Public Information Request process, where I requested access to all public documents pertaining to The Texas Writing Pilot. The documents reviewed consisted of the following:

- Texas House Bill 1164 (2015)

- Request for Applications: The 2016-2018 Writing Pilot Program (Texas Education Agency, 2016a)

- SEC 39.02301 Writing Assessment Study Pilot Program (Texas Education Agency, 2016b)

- House Bill 1164 Writing Pilot Program Report to the Governor and the Texas Legislature, 2016 (Texas Education Agency, 2016c)

- House Bill 1164 Writing Pilot Program Report to the Governor and the Texas Legislature, 2017 (Texas Education Agency, 2017)

- House Bill 1164 Writing Pilot Program Report to the Governor and the Texas Legislature, 2018 (Texas Education Agency, 2018)

- The Texas Writing Pilot Program (Texas Education Agency, n.d.)

Data Collection and Analysis

Archival data related to The Texas Writing Pilot, were analyzed using a document analysis protocol (Gibson & Brown, 2009). With any task of document analysis, objectivity and sensitivity must remain of utmost importance in order for the document analysis to be seen as credible and valid (Bowen, 2009). Because of my prior involvement with the pilot, I wanted to make sure I kept an open mind about the data. For

this reason, an open coding protocol (Creswell, 2013) was used to examine each line of text by considering the subject and key ideas within each line of text. For each line of text reviewed, I asked myself what was the key theme(s) or idea(s) of the line and then I would record my findings in a spreadsheet where I collected the findings of each of the reviewed documents. This analysis provided the opportunity for themes to emerge. From there I was able to conduct further analysis to identify categories. The process included careful reading and re-reading of the data (Bowen, 2009). During this phase, member checking occurred with others who had previously taken part in the pilot to ensure the accuracy of the coding (Creswell, 2013).

Findings

Data analysis led to the development of 29 identified codes. From additional analysis of the codes, six categories emerged: Authentic Assessment, Writing Portfolios, Student Growth, Instruction, Scoring: Rater Agreement, and Training. The following table provides a look into each category, the codes that were associated with category, along with excerpts of supportive data shedding insight into what happened during the pilot. Following Table 2, further explanation and elaboration will be given to each of the categories.

Table 2

Identified categories from data analysis

Category	Codes	Examples of Supportive Data
Authentic Assessment	Alternative assessment, authentic assessment, collaborative assessment design	"The Texas Writing Pilot provided the opportunity to begin an investigation into alternative forms of writing assessment in the state" (Texas Education Agency, 2018, p. 2).
Writing Portfolios	Writing process, writing portfolio	"Overview The Texas Writing Pilot was structured to study a more robust, portfolio-style writing assessment, to meaningfully integrate summative assessment into daily instruction" (Texas Education Agency, 2018, p. 2).
Student Growth	Student growth, student engagement	"These responses will be used to assess the student's mastery of the essential knowledge and skills in writing through timed writing samples, and improvement of a student's writing skills from the beginning of the school year to the end of the school year, as required by the legislation" (Texas Education Agency, 2016b, p. 1).
Instruction	Classroom instruction, feedback	"Ultimately, a well-designed assessment should inform and aid best practices in instruction" (Texas Education Agency, 2017, p. 9).
Scoring: Rater Agreement	Reliable and valid, scalability, accountability	"No individual or sum of ratings in the current study reached the reliability of 0.80, and most of the scores' reliabilities were far below 0.80" (Texas Education Agency, 2018, p. 6).
Training	Teacher training, collection training, scoring training	"While there were some sporadic highlights across the population in both Year 1 and Year 2, the overwhelming variance in data suggests that training enough educators to be standardized scorers would not be possible" (Texas Education Agency, 2018 p. 6).

Authentic Assessment

Within the category of authentic assessment, I was able to see that much attention was given in the documents to the idea that the development of the pilot would be "a collaborative design process, inclusive of teachers, Education Service Centers (ESCs), and institutes of higher education" (Texas Education Agency, 2016a, p. 2) in order to develop a meaningful assessment for students. It should also be noted that the pilot design stated there would be three process writing papers, while the actual pilot only sampled and assessed two. Additionally, the manner in which the writing assignments would be evaluated was addressed as a means to assess students growth over the course of a school year. During an initial face-to-face meeting in 2017, participants collaborated to develop "the foundation of the writing pilot rubric." (Texas Education Agency, 2017). However, year one participants viewed the rubric too similar to the STAAR rubric (Texas Education Agency, 2018) and pushed for a new rubric to be created shifting the pilot rubric from a 4-point holistic rubric to a 6- or 3-point analytic rubric.

Writing Portfolios

For the category of writing portfolios, I found statements that described the importance of the writing process within the creation of the students' writing portfolio. For the pilot, while teachers were provided with designated time frames and submission windows for assigning and collecting each of the writing-process samples the teachers had the flexibility to select the genre of writing to collect from students. This allowed teachers to "fully align the assessment with local instruction and scope and sequence of

69

curriculum" (Texas Education Agency, 2018, p. 7). Ultimately the design of the pilot was that the assessments would miorror classroom instruction and assess students growth over the course of a school year.

Student Growth

In the category of student growth, student mastery and growth over the course of a school year was revealed as a key area of emphasis. As the program design documents specified "These responses will be used to assess the student's mastery of the essential knowledge and skills in writing through timed writing samples, and improvement of a student's writing skills from the beginning of the school year to the end of the school year, as required by the legislation" (Texas Education Agency, 2016b, p. 4). Additionally, teachers reported students showing stronger engagement with their writing assignments as a result of the pilot (Texas Education Agency, 2018).

Instruction

For the category of instruction, the 2018 Report included reflections of teachers who felt that their writing instruction was more intentional and focused because of the pilot. While the study conducted by the TEA did not evaluate alignment between instruction and assessment, teachers reported stronger congruence. Teachers felt that even the prompts were more authentic because it was a direct extension of what they were already doing in class.

Scoring: Rater Agreement

One thing that became apparent in the scoring category was that the state did not find the pilot to be a viable assessment option because "scoring correlations and rater-agreement never reached the same level as STAAR, at scale" (Texas Education Agency, 2018, p. 2). This may be due in part to "limited appropriates to the project reduced the ability for true piloting of a standardized assessment prototype" (Texas Education Agency, 2018, p. 2).

Training

In many of the Program Design documents, consideration was given to the training that would be required for teachers to successfully be equipped to implement the pilot as well as score student writing on a new rubric. According to the 2017 Report, "TEA and ETS then facilitated a virtual train-the-trainer session for the three regional ESC representatives who, in turn, held in-person scoring trainings for participating teachers in their region" (Texas Education Agency, 2017 p. 2).

Discussion

By conducting a document analysis on the documents relating to The Texas Writing Pilot, I was able to see trends that were helpful in shedding light on my question of what happened during the pilot. While the data was revealing, information not included in the data was also revealing. Findings from the document analysis have brought to light timely insights for any teacher of writing not only regarding the ways in

which we utilize assessment but also the ways in which we can use our voice to advocate for supports that we need in order to more effectively teach writing.

Assessment

According to some, because the interrater reliability never reached the same level of STAAR (Texas Education Agency, 2018), there is question as to whether teachers would ever be able to truly score their own student writing in a statewide assessment for accountability purposes. But when reviewing the data, one is able to see that the rubrics changed from Year 1 to Year 2, there was a lack of funding for training and support, as well as inconsistent training and support for the teachers and administrators across the state. Additionally, district administrators complained that the timelines for the pilot did not correlate with other district demands and responsibilities. Any one of these variables alone could dramatically impact the results of interrater reliability. While teachers reported that the assessment was authentic and more congruent with classroom instruction, TEA found the consistency in scoring to be lacking. Supporters of authentic assessment should continue to advocate for other studies and opportunities that support core instruction through the use of authentic assessments but should also focus on the consistent implementation within the study.

Instruction

It was clear that through the pilot teachers felt improvement in the quality of their writing instruction. The intention of the pilot was that the writing assessments would be an extension of everyday classroom instruction. Teachers were also able to see a tighter

alignment between what is taught and what is assessed because the assessment was a direct extension of what students were learning. In Year 1, participants of the pilot were not required to take the STAAR writing assessment in Grade 4 and 7. Teachers reported that this reprieve from the state assessment, provided them more leeway in teaching. As a result, teachers felt the quality of instruction was improved because they were not overly teaching to one specific test or style of writing. However, in Year 2, with state accountability scores looming, TEA required many districts participating in the pilot to assess their students in Grade 4 and 7 with the STAAR writing assessment in addition to The Texas Writing Pilot. Teachers expressed concerns that the way they needed to teach students for the pilot verses the constrained 26 lines for STAAR made it feel as though there were competing forces at play. As the state continues to consider what the new iteration of STAAR for reading language arts looks like, teachers need to continue to advocate for meaningful assessments that align tightly with classroom instruction as well as current beliefs and practices about writing instruction.

Scoring

Two of the four pieces of writing students turned in were timed writing samples where they had to select from a list previously released STAAR prompts. This provided a relatively controlled field of responses for the graders but did not match the types of writing happening in the classroom. The other two pieces the students wrote were the process pieces where either the teacher or the student chose what the student wrote about. The wide variety of teacher expectations for the process piece contributed to a mismatch

between the two different types of writing and made it difficult for graders to grade the two different types of writing. While the sample scoring papers for the teacher calibration training included a variety of types of sample papers (Texas Education Agency, 2018), it was evident that teachers needed more support and training at the beginning of the pilot about what a quality assignment for a process paper might look like. In most cases, teachers did not receive information on how students will be scored until after the first writing sample was collected and turned in. Just as we do not want students to feel as though standardized assessment is something being done to them, we also need to take intentional steps to ensure that educators understand the language of the assessments, calibrating with them along the way (and often) about what makes a quality piece of writing.

Calibration

A recurring theme of the reports released by the TEA revealed there was no significant interrater reliability between teachers and the ETS raters. Ultimately this is problematic if the state were ever to see the value in having teachers score their own students' essays for assessment purposes. For those of us who were involved in the pilot, we know though there was insufficient time and support for teachers to calibrate. Teachers need to advocate for the state to better articulate the expectations of quality writing through additional support and detailed documents that better explain and demonstrate what makes quality writing. The lack of interrater reliability does not definitively demonstrate that teachers' cannot rate well, but rather it demonstrates that

teachers do not have agreement or understanding about how the state expects the scoring rubric should be utilized. A remedy may be for teachers to advocate for time to work together in calibration meetings in order to have opportunities for meaningful discourse about writing.

Conclusion

As I look back upon everything I know now about The Texas Writing Pilot, I have new insight – not only about what happened during the pilot – but also an insight into the important role educators and administrators alike can play as experts from the field with influential voices advocating for the needs of their students. The data revealed a clear disconnect between how educators teach writing and how the state assesses writing. These dueling forces are ultimately what lead to the breakdown between assessment, instruction, scoring, and calibration that I presented in the discussion. As the state continues to consider assessment implications because of our new state standards, as well as new and alternative ways to authentically assess what a student knows, educators and administrators must take an active role contributing their voices to the process.

By taking part in studies such as The Texas Writing Pilot teachers were able to have their voice heard and specifically influenced the design of the assessment. Teachers can and should play a role of advocacy by finding opportunities to articulate the support structures needed from the state to better teach and support student mastery of the standards. Even at a district level, teachers can advocate for additional opportunities to

collaborate and learn alongside one another to calibrate and more tightly align scoring even as a team.

While there is much for teachers and administrators to learn from this pilot, there are takeaways for the state as well. In reviewing the implementation of the pilot, the state should consider how they design and execute pilot studies. Decisions such as changing the rubric from Year 1 to Year 2, as well as adding 25,000 students to a study in the last 5 months of the study, burdened the entire study and calls into question its validity. The study did in fact provided provide both quantitative and qualitative data regarding writing assessment, but the study was not consistently executed. While the qualitative data provided positive insights from educators about the potential benefits to students and classroom instruction, the quantitative data should be carefully considered before it is used for decision-making purposes about how such an assessment could be implemented, or not implemented, across the state.

Practically speaking, the state also needs to provide support for teachers by way of explanation guides and documents that help them better understand the state's interpretation of standards (the TEKS) for assessment purposes so teachers can better understand how to teach and assess the standards within the classroom. Specifically, it would not be helpful for these documents to merely show a definition of each individual standard and how it is assessed, rather they need to demonstrate how each of the language arts standards are interdependent to all of the other standards and recursive over time.

While this document analysis sought to investigate the state's implementation of H.B. 1164 (2015) regarding student performance assessment, its findings revealed potential opportunities for educators to take action and play a pivotal role in creating practice and policy across the state. Ultimately, the moral of this research is when given the opportunity to attend a meeting, say yes – even when you aren't quite sure where it will take you. We know that the new Reading and Language Arts STAAR test, that will align to our new state standards, will include writing at all of the tested grade levels (third grade - English II), so I implore each of you to seek ways in which you can be influential in ensuring our students have the opportunity to demonstrate what they know through authentic and meaningful assessments.

Influences of Teacher's Scoring Decisions When Evaluating Student Writing, Assessing Writing

History often describes educational policy and practice as a pendulum swinging from one extreme idea to the next. While standardized assessments have been part of the educational landscape since the late 20th century, the pendulum of state accountability took a hard swing in 2002 when President Bush reauthorized the Elementary and Secondary Education Act by signing The NCLB of 2001. The goal of this school reform measure required testing of third through eighth grade students in the areas of literacy and math as a way to hold states and schools accountable for closing the achievement gap (Stiggins, 2002). States had to respond to this legislation and submit to the federal government new plans for state assessment and accountability. Since that time, states

have pursued various ways to adequately assess students to hold schools and districts accountable for their efforts in closing the achievement gap (Darling-Hammond & Adamson, 2010).

It is undeniable that the education system is currently in an age of assessment and accountability (Stecher, 2010). Yet even in this age of accountability, it is necessary for states to consider assessments that to not only meet federal and state accountability mandates but also provide teachers with valuable instructional data. In 2015, the State of Texas began looking at the possibility of utilizing a localized writing assessment for the purposes of statewide high-stakes testing. Known as The Texas Writing Pilot, one of the many goals of this alternative assessment was to provide students with more timely feedback on their writing. The design of the pilot was to empower the student's classroom teacher to assess the student writing and provide this timely feedback, as opposed to sending it off for a blind scorer to rate. Ideally, students would then be able to use this specific feedback as they continued to develop their writing skills throughout the year. In addition to organizing the pilot, the TEA conducted a feasibility study alongside the pilot to determine score reliability by evaluating "the quality of locally-produced ratings and whether stakes can be associated with the locally-produced ratings" (Texas Education Agency, 2017, p. 5).

From a quantitative research stance, and much like the study conducted by the TEA the determination of feasibility for this type of alternative writing assessment rests squarely with the establishment of inter-rater reliability across raters, yet some

researchers suggest that variability is just one factor for consideration (Huot et al., 2010; Jeong, 2015a; Yancey, 1999; Zhang, 2016) in the rating process. For a more complete picture, factors such as a teacher's experience and background must also be considered (Jeong, 2015a).

Purpose of the Study

The purpose of this study was to explore the scoring decisions teachers make when evaluating student writing and to consider the various factors at play when these decisions are made. This study will contribute to the body of knowledge related to teacher assessment of student writing and will add to the body of knowledge related to statewide writing assessments. Previous studies have looked at the development and implementation of a statewide writing performance assessment but have only looked at the teacher's ability to evaluate student-writing samples using a rubric in terms of quantitative inter-rater reliability data. While inter-rater reliability is an important feature to consider when determining the scale-ability of teachers evaluating student writing, further inquiry is needed in considering how a teacher's knowledge and training (Birgin & Baki, 2007) contribute to her scoring approach. By building upon the work of Wolfe and Feltovich (1994), related to scorer cognition, this study considered the knowledge and professional experiences that contribute to the different approaches individual teachers use when scoring student writing samples.

The aim of this study was not to generalize the findings to a given population, but rather consider the framework of scorer cognition established by Wolfe and Feltovich

(1994) to inform future work related to alternative assessment design for the purposes of high-stakes testing. It is hoped that as a result of this study, further light may be shed on features of professional experiences that influence teachers' scoring decisions and may subsequently lead to a framework for training teachers as the state looks to scale an alternative form of writing assessment.

The paper is set out as follows. The literature review begins with a brief discussion of key topics and ideas including assessment and accountability, teacher decision-making, and professional learning. Following the review of literature, the method and approach to data analysis used in the study are reported. An analysis of the data and findings of the study are then presented and discussed, with an emphasis on insights gained in terms of scorer cognition. Implications of the analysis are discussed, and the limitations and delimitations are outlined.

Literature Review

This qualitative study (Merriam, 2009) examined the various factors that come into play as a teacher makes scoring decisions when evaluating student writing. The context of this study looks closely at six writing teachers, with a diverse range of experiences, who scored student writing samples while using a think-aloud protocol with and without a standardized rubric. This section looks at the definition and previous work regarding assessment reliability and validity and then considers the role rubrics play from the perspective of reliability and validity. Finally, further attention is given to teacher decision-making and the role professional learning plays within teacher decision-making.

Assessment Reliability

As far as testing is concerned, reliability can be defined as the frequency to which scores from an assessment would be expected to be similar across multiple iterations of the same assessment (Huot et al., 2010; Lemann, 2000; Moss 1994). Whereas instrument reliability refers to a test's ability to produce consistent scores, inter-rater reliability refers to the agreement between raters on the same papers for a given assessment (Huot et al., 2010). One consideration for the widespread adoption of assessments, and more specifically, standardized assessments, can be attributed to a perceived sense of the reliability of the assessment (Moss, 1994).

In the latter half of the twentieth century, Charles Cooper (1977) examined the effectiveness of scoring, which focused on the general impact of a piece of writing, to rank students' writing. As a way of validating the scoring process, he identified seven types of holistic evaluation including general impression marking, formative response, and analytic scoring (Cooper, 1977). Based on his work, Cooper (1977) believed it was possible to improve reliability to acceptable levels when raters shared similar backgrounds and were carefully trained. Training for raters may include not only a focus on rater practices but also clear explanations for each criterion within the rubric as well as examples for each descriptor (Jeong, 2015b). Later studies revealed that by making teachers aware of scoring inconsistencies teachers begin to adjust their scoring and in turn, their scores become more reliable (Coffman, 1971). "While scoring reliability in

writing assessment is undeniably important, it has been equally difficult to deliver" (Huot et al., 2010, p. 500).

Assessment Validity

Prior to the 1920s, it was as if validity was something taken for granted (Huot et al., 2010). Because of the simple fact that creators of assessment instruments were experts of their assessment instruments, it seemed safe to assume that they were also experts on the validity of that assessment as well (Diederich et al., 1961; Huot, 2002). Compounding the lack of clarity regarding validity, the overemphasis on rater agreement, blurred the difference between validity and reliability (Behizadeh & Engelhard, 2011; Huot et al., 2010; Wiggins, 1993). It wasn't until 1954 that validity began to be looked at in broader terms than just rater agreement. By 1966 "content, criterion, and construct validity became the three main foci for test validation" (Huot et al., 2010, p. 505).

By the 1980s, some researchers explained the over-reliance on reliability had been because of creating a viable assessment as opposed to defining a theoretical framework for writing assessment (Behizadeh & Engelhard, 2011; Huot, 2002). By working within and against the prevailing psychometric paradigm, researchers such as Edward White decided to confront the issue of validity with writing assessments and set out to "devise a writing test that could meet the standard stipulated by the testing experts" (Yancey, 1999, p. 490). Adapting the then widely accepted testing technology, the newly designed assessment focused on an end of year essay test that was based on the curriculum covered over the course of the year. Three key procedures identified for this assessment including

a writing prompt, anchor papers and scoring guides for raters, and a determination of acceptable agreement by the test-makers helped to distinguish this assessment from other assessments in the field (Yancey, 1999).

Rubrics

Although multiple-choice assessments can provide efficient and timely results (Williamson, 1994), performance assessments provide students with multiple ways to show what they know (Stiggins, 1987; Tung, 2010). A carefully crafted performance assessment allows students opportunities for choice as they demonstrate a range of abilities and skills (Tierney et al., 1991; Tung, 2010). Through this type of assessment, students construct or create original responses as opposed to the constrained response they are expected to provide on a multiple-choice type of assessment. Typically for performance assessments, evaluation is based on multiple criteria often in the form of a rubric (Gomez, 1999; Wiggins, 1989).

In many cases, performance assessments rely on teachers as they use rubrics or rating scales to evaluate a student's writing portfolio (Valdéz Pierce & O'Malley, 1992). Rubrics serve as a qualitative evaluation, based on multiple criteria (Gomez, 1999; Wiggins, 1989), that define standards and expectations, as well as important aspects of performance, by describing the degrees of quality (e.g., from above expectations to below expectations) for a given task (Jeong, 2015b; Jonsson & Svingby, 2007; Saddler & Andrade, 2014). Rubrics can be task-specific or generic in nature (Jonsson & Svingby, 2007). In their review of 40 research articles regarding rubrics for writing assessment,

Jonsson and Svingby (2007) reported that rubrics were found to enhance scoring reliability, support the sound judgment of performance assessments, and provide opportunities for teachers to promote learning and improve instruction. Typically, rubrics are utilized as either a holistic or an analytic scoring evaluation tool. When scoring holistically, the rater must determine the overall quality of the performance or piece, whereas when scoring analytically the rater must assign a score to each domain of the rubric (Jonsson & Svingby, 2007).

      Holistic writing rubrics. A holistic rubric is founded on the premise that a valid writing assessment is dependent upon the piece as a whole and cannot be viewed merely as a collection of pieces and parts (Huot, 1990; Lloyd-Jones, 1977) While holistic rubrics can aid evaluators in moving away of from a corrective or punitive assessment of writing, wherein the assessor over emphasizes the mechanics and grammar of the piece, holistic scoring is limited in providing the specificity of what makes the writing excellent or poor (Dempsey et al., 2009; Huot, 1990). "A legitimate question about the possible obtrusive nature of holistic scoring revolved around the ability of a reader to actually see all of a student essay and not just the parts which have relevance to the scoring guidelines she and her fellow raters are using" (Huot, 1990 p. 210). Teachers often find holistic scoring difficult to use as a tool for providing feedback to students on specific areas of continued growth. Without providing specific areas of growth, students often see a holistic score as arbitrary or subjective (Dempsey et al., 2009; Lumley, 2002). While the reason may be attributed to ease and cost efficiency, holistic scoring is usually the preferred method for

84

large-scale assessment (e.g., state and national assessments; Dempsey et al., 2009; Huot, 1990; Lumley, 2002).

Analytic writing rubrics. In contrast to a holistic rubric, an analytic rubric has the ability to provide more specificity and as such serves as a better diagnostic tool for evaluation within the classroom (Knoch, 2009). As an early developer of analytic approaches to writing assessment, Diederich (1974) conducted a study of academic and business professionals as they scored sets of writing papers, on a 9–point scale, without any training or rubric. Diederich (1974) and his colleagues then used the data collected to tease out recurring themes that they then used to create an analytic framework for writing assessment. Stemming from this work, in an analytic rubric the important criteria, or features, of the assignment are described in detail, including specificity as to what the assignment looks like, at the high score point, mid score point, and low point score point of the rubric (Cooper, 1977; Dempsey et al., 2009). A score is then given for each domain of the rubric. The specificity provided within the analytic rubric provides an opportunity for teachers to engage students with meaningful feedback as a vehicle for promoting continued learning (Jonsson & Svingby, 2007).

In a 2002 study of raters using an analytic rating scale for scoring ESL students applying to immigrate to Australia, Lumley (2002) wanted to understand how individuals applied features of a rating scale to writing samples. Based on an analysis of a think-aloud protocol conducted by each participant, the data showed that although raters often follow a similar rating process, the scoring decisions made were quite different (Lumley,

2002). While raters attempt to remain close to the language of the scale, the scale is unable to encompass all possibilities within a piece of writing. Although some tension exists with raters between their overall impression of a piece of writing and the specific language of the scale, Lumley (2002) acknowledged it is possible to find consistency in scoring between raters if adequate training and support is provided to the raters.

  Scoring using rubrics. Although rubrics are widely used among teachers, the rubrics themselves are often called into question with arguments that they contain inconsistent or vague language within each of the descriptors (Jeong, 2015a). The processes teachers use for rating assessments with rubrics are often criticized as well for being too subjective (Lumley, 2002). As a rater, one must decide, consciously or otherwise, which feature of the scale to pay attention to as well as how to define and distinguish between the language within the scale of the rubric (Lumley, 2002). Even when raters make an intentional effort to follow the rubric as written, studies show that invariably the rater's decision-making process can be influenced by the overall impression of the piece of writing, and even personal intuition (Jeong, 2015a). Some studies suggest that rater reliability, in part, is attributed to the training raters receive (Jeong, 2015a; Knoch et al., 2007). One such study conducted by Jeong (2015b) closely examined experienced teachers' scoring of essays with and without a rubric and compared the data to similar data about novice teachers scoring. Findings from the study indicated that experienced raters were better able to put their personal constructs aside to

rely on the rubric for scoring as well as reveal a necessity for rater scoring training that includes detailed descriptor explanations and examples (Jeong, 2015a).

As the validity and reliability of performance assessments continued to be called into question, it appears that rubrics provide a means for supporting validity without sacrificing reliability (Jonsson & Svingby, 2007). Furthermore, it is through their intentional use that teachers can engage in meaningful dialogue with students to promote continued growth as a student.

Teacher Decision-Making

At its core, "any teaching act is a result of a decision, either conscious or unconscious" (Shavelson, 1973, p. 144). While teacher decision-making is a complex balancing act of negotiating one's own beliefs, curricular constraints, and institutional constraints, the most foundational skill in teaching is just that – decision-making (Borko et al., 1981; McMillan, 2005; Shavelson, 1973). Studies that have focused on instruction have often indicated that experienced teachers are more adept at interpreting, evaluating, and explaining the complexities of classroom instruction, whereas inexperienced teachers rely on more basic or concrete explanations and interpretations of classroom events (McMillan, 2005). While many teacher decisions, especially as they relate to instruction, must be made in the moment, assessment decisions often provide the affordance of time, allowing the teacher to access and consider a wealth of information as part of the decision-making process (Borko et al., 1981).

Internal factors. Internal factors for decision-making are influenced by an individual's core beliefs and values about children, education, and learning (Gill & Hoffman, 2009; McMillan, 2005; Putnam & Borko, 1997). These beliefs can directly influence our choices, efforts, performance, and even behavior. Beliefs can also develop by one's teaching and learning experiences (Gill & Hoffman, 2009). These beliefs often strongly shape both the instructional and classroom management practices an educator uses in her classroom (Putnam & Borko, 1997), and yet teachers often cannot clearly articulate the beliefs that influence their work (Gill & Hoffman, 2009).

Although teachers' beliefs during the planning process can be hard to measure, by exploring teacher talk during shared planning, Gill and Hoffman (2009) sought to gain insight into teachers' beliefs about teaching and learning by focusing on the rationales behind the teachers' decision-making. Their study consisted of four middle school teachers who were observed once a week in the fall semester as they planned, shared lessons, and discussed other pertinent things as a team (Gill & Hoffman, 2009). Ultimately, it was this shared planning that provided insight on the teachers' beliefs about "pedagogical content, general pedagogical beliefs, beliefs about curriculum and textbooks, and beliefs about students' ability (Gill & Hoffman, 2009, p. 1253).

External factors. While a teacher's beliefs guide the internal factors of her decision-making, there is a myriad of external forces that may come into play as teachers make decisions. State-mandated assessment, school/district policies, and parents or home life are just some of the forces that influence teachers as they make decisions on a daily

basis in their classroom (Clark & Peterson, 1984; Griffith et al., 2013; McMillan, 2005).

For example, teachers have to weigh the school or district's policies around curriculum,

classroom management, or social promotion as they designs lessons, redirects

misbehavior, and supports struggling students. A teacher also has to consider the manner

in which her students and school will be evaluated. With the external pressures of state-

mandated assessment, teachers often feel forced to use more objective assessment

practices throughout the course of the school year that mirrors the state test to ensure that

her students will be ready for the assessment (McMillan, 2005).

Guided by a widespread assumption that all teachers do now teach to the test, in

her review of studies regarding state-mandated testing, Cimbricz (2003) analyzed the

relationship between state assessment and teachers' beliefs and practice. Of the studies

reviewed, Cimbricz (2003) found that state testing does in fact, influence how teacher

behaviors and the decisions a teacher makes. However, although state testing does

influence teacher practice and their views of learning, there was not a consistent pattern

of influence because other variables such as teacher status or experience level contributed

to the ways in which an individual teacher interpreted the importance of the state

assessment (2003). As educators engage with the myriad of decisions to be made during a

day they must carefully weigh these competing forces (Black & Wiliam, 1998;

McMillan, 2005).

Professional Learning

In order for teachers to facilitate students to deeper levels of conceptual understanding (Borko, 2004), teachers must have a vast understanding of the discipline for which they teach as well as the research that backs up the practice (Stokes, 2010). Most commonly, opportunities for teachers to develop and deepen their understanding of a given discipline come in the form of professional learning. However, it is no secret that teachers often have a less than enthusiastic outlook when it comes to professional development. In order to ensure that professional learning is both meaningful and relevant, the teachers who participate, the facilitator who guides the learning, and the situation or context for which the professional learning is given must be considered (Borko, 2004).

While exploring the various factors that come into play as a teacher makes scoring decisions when evaluating student writing, it is important to consider the learning experiences of the teacher that potentially influences her scoring decisions. It is also important to consider the way individuals situate themselves within a greater social and cultural context to participate in the interdependent process of constructing knowledge (Crotty, 1998; Papert, 2011; Vygotsky, 1980). An educator, attending a required professional learning experience is not a sponge waiting to soak up knowledge from the facilitator. Passive consumption of the information presented is one thing, but in order to actively construct knowledge, a fundamental consideration must be the social interaction between the learner and others participating in the learning experience (Bruner, 1990;

Cunningham & Duffy, 1996; Vygotsky, 1980). In other words, the way an experience is understood is in direct concert with the manner in which the individual constructs the context of that situation while under the influence of social factors (Crotty, 1998; Flick, 2009; Papert, 2011; Rogoff, 1994; Vygotsky, 1980).

For a child, learning experiences often focus on the acquisition of skill, but for adult learning, experiences build upon individual's prior knowledge and skill in order to provide opportunities that promote self-directed learning (Knowles, 1970, 1996; Merriam, 2001; Merriam et al., 2012; Zmeyov, 1998). In the case of adult learning, learning moves to a more problem-centered approach where teachers are empowered to actively participate in shared inquiry and advocate for their own educational needs (Zmeyov, 1998). With a focus on continuous improvement, the learning is action-oriented and focuses on ensuring high levels of learning for all students (DuFour & DuFour, 2013). Through the collective practice of engaging in discourse and critically scoring the performance assessments within a greater community of teachers, teachers developed a stronger sense of knowledge that was expected for the writing assessment.

Methods

In order to further conversation regarding alternative assessment options for student writing assessment across the state of Texas, within the context of high-stakes testing and accountability, this research study focused on the approaches individual teachers take when scoring writing. Specifically, the research questions included:

1. What are the various factors that come into play as a teacher makes scoring decisions when evaluating student writing?

    a. How do teachers approach the application of a rubric in evaluating student writing?

    b. What metacognitive strategies do teachers use in the process of making scoring decisions when evaluating student writing?

Research Design

In an effort to explore the various approaches used when scoring student writing, a qualitative study (Merriam, 2009) design was used as it provides an opportunity to probe into the various ways meaning is constructed. Guided by the research questions (Stake, 1995), the examination of diverse perspectives provided an increased depth of understanding of the central phenomenon to scoring student writing. The design of the study also allowed for the collection of various sources of data such as surveys, interviews, and observational data (Stake, 1995). The nature of the study, as well as the multiple teachers, support the identification of different approaches individuals use when scoring student writing through an analysis of themes.

This research study was implemented in three phases. Phase I included enrollment of the participants through purposeful sampling (Merriam, 1998) and an initial open-ended participant survey (see Appendix B). Phase II involved data collection through interviews and interaction with the participants selected. Semi-structured, face-to-face interviews with participants (see Appendix C), a think-aloud protocol utilized without a

rubric (see Appendix D) and then a think-aloud protocol (see Appendix E) with an

analytic rubric (see Appendix F) as participants score writing samples (see Appendix G,

Appendix H, and Appendix I), and a follow-up semi-structured interview (see Appendix

J) was conducted (Merriam, 1998). Phase III required analysis of participant responses

using coding, validation of transcribed information through member checking, and the

identification of emerging themes (Creswell, 2013; Stake, 1995).

Role of the Researcher

As a certified Texas public school educator for over 13 years, I have a deep

interest in the ways in which we assess students. While in my experience I have had the

privilege to teach middle school students, I have also had the opportunity to coach and

lead pre-kindergarten through 12th grade teachers in the field of literacy. It was in my

role as a district administrator for English Language Arts that I became involved with

The Texas Writing Pilot. Initially, I was working with several other educational leaders

from around the state to develop a portfolio assessment plan that districts could utilize

locally. Then we turned our efforts to assist the state to develop what would become to be

known as The Texas Writing Pilot. During the time of the 2-year pilot with the state, I

became the regional liaison for the pilot between TEA and the participating districts from

a region in North Texas near or around Dallas Fort Worth. In my role as liaison I

participated in a number of calls and meetings with TEA, sought feedback from educators

regarding the direction of the pilot, created pilot design specifications as well as the pilot

rubric, and provided scoring training to teachers participating in the pilot.

As for my role as the researcher within this study, my role was to observe, interview, collect, and analyze data. Because of my involvement with The Texas Writing Pilot, my study necessitated bracketing (Creswell, 2013) throughout the study. Specifically, when conducting the semi-structured interviews and think-alouds with the participants, I refrained from adding comments or asking questions that were not within the scope of the protocols.

Setting

This study took place in the state of Texas to explore the efforts being done by the state in the area of alternative writing assessment. Texas is as diverse ethnically and socio-economically as it is geographically. Second, in total land area only to Alaska, Texas served just short of 5.5 million students during the 2017–2018 school year. From 2007 to 2017, a 10-year period, student enrolment in Texas public schools increased by 15.6 %. Of the 5,399,682 students enrolled in Texas schools for the 2017–2018 school year, 52.4% identified as Hispanic, 27.9% as White, 12.6% as African American, 4.4% as Asian, and 2.3% as multiracial. Additionally, the percentage of students identified as economically disadvantage rose from 55.2% in 2007–2008 to 58.7% in 2017–2018. There were more than 8,900 schools in 1,200 school districts and open-enrollment charter schools employee more than 356,000 teachers (Division of Research and Analysis Office of Academics, 2018) to serve those students in 2017–2018.

Participants

Participants for this study included Texas writing teachers from the North Texas area near or around Dallas Fort Worth who have experience teaching writing at a grade level where writing is assessed by the state (i.e., fourth grade, seventh grade, and ninth grade). Purposeful sampling (Merriam, 1998) was used to select six participants selected for this study. Three teachers, one from each of the three grade levels (fourth grade, seventh grade, and ninth grade), had more than five years of experience teaching and three teachers, one from each of the three grade levels (fourth grade, seventh grade, and ninth grade), had fewer than five years of experience teaching. Table 3 shows the participants, grade levels taught, and years of experience teaching.

Table 3

Experience of Study Participants

Participants	Grade Level Taught	Years of Experience
Participant 1	9th Grade	More than 5 years
Participant 2	9th Grade	Less than 5 years
Participant 3	7th Grade	More than 5 years
Participant 4	7th Grade	Less than 5 years
Participant 5	4th Grade	More than 5 years
Participant 6	4th Grade	Less than 5 years

This range experience among the teachers provided insight on the influence these teachers draw upon when making scoring decisions. In addition to the range of years of

experience, each of the teachers had a diverse background of experiences and

professional learning experiences that contributed to their process for scoring student

writing.

Data Collection

Multiple research data sources were used during data collection to provide a

focused yet comprehensive view of the various factors that come into play as a teacher

makes scoring decisions when evaluating student writing samples. The primary unit of

analysis was teachers' verbal and written statements that reveal their processes of

pedagogical reasoning and action (Shulman, 1987) by investigating the different

approaches to individuals take to scoring student writing. To capture and describe these

approaches, a pre-scoring participant survey, initial interview, two think-aloud protocols

(Ericsson & Simon, 1984; Smagorinsky, 1994), one without and one with a standardized

rubric was utilized. Figure 4 visually represents the data collection process. Recorded

field notes from observations of participants' behavior as well as interactions during

think-alouds were recorded to provide additional insight into what the teachers are

considering when they score. To triangulate the data (Marshall & Rossman, 2014), post-

scoring interviews were conducted and reflections from the participants were gathered.
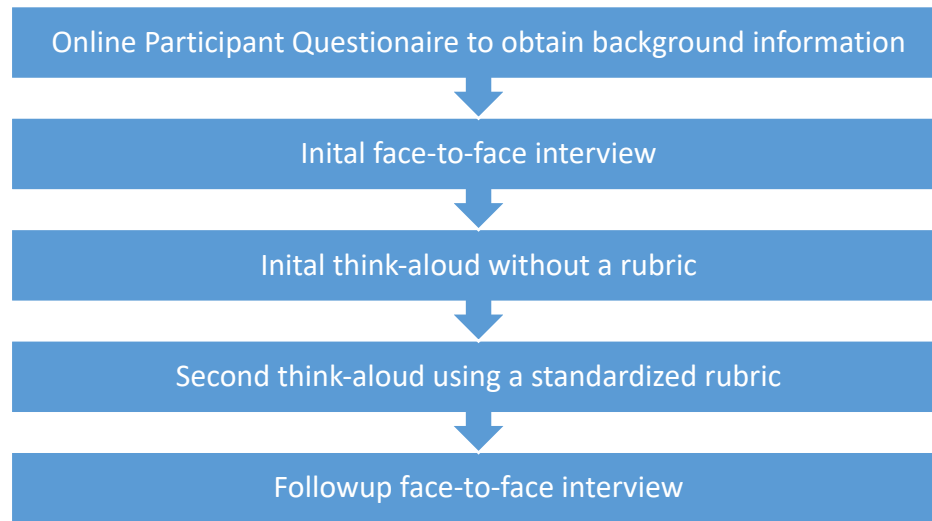
Figure 4. Data Collection Process.

An open-ended participant survey was developed for the purposes of gathering background information from teachers regarding their experiences and specialized training in both scoring and writing instruction. Emerging patterns of teacher knowledge and training were explored through the analysis of the teacher's open-ended responses from the interviews.

Teachers participated in an initial semi-structured interview (Merriam, 1998), to capture the beliefs and opinions of the teacher as it relates to the scoring and instruction of writing. The eight-question interview protocol, along with possible probing questions that were used as necessary, took place in a location determined by the participant. Generally, each interview lasted 30 – 40 minutes. The use of the semi-structured interview (Merriam, 1998) protocol supported the consistency of the interviews to ensure sufficient information was collected (Marshall & Rossman, 2014).

At the same time as the initial interview, the teacher scored student writing while thinking aloud about scoring decisions into an audio recorder, capturing those decisions and how they were made. This think-aloud protocol (Ericsson & Simon, 1984; Smagorinsky, 1994) was used to monitor the use of metacognitive strategies as teachers engage in scoring student-writing samples. For this first think aloud activity, teachers were not provided with a rubric. The intention was that educators might speak more freely about their scoring processes if unencumbered by a rubric. Recordings of the think-aloud were transcribed and coded.

Following the initial scoring exercise, a secondary scoring exercise was conducted. Teachers were introduced an analytic writing rubric, given time to study the document, and then once again asked to score student work while thinking aloud about scoring decisions into an audio recorder -- this time using the standardized rubric. Capturing those scoring decisions and how they were made were used for the purposes of considering the various factors that come into play as a teacher makes scoring decisions when evaluating student writing.

When scoring with a rubric, participants used the analytic scoring rubric of The Texas Writing Pilot (see Appendix A) that was collaboratively developed by teachers, employees at the regional educational service centers, and in collaboration with institutes of higher learning (Texas Education Agency, 2016c). This rubric consists of four scoring domains: organization, content, language, and conventions. With the exception of conventions each of the other three domains consist of a six-point scale that ranges from

Very Limited (1), Limited (2), Basic (3), Satisfactory (4), Accomplished (5), Exceptional (6). The conventions domain is based on a three-point scale. Within each domain, an explanation of each score point further explains the criteria and expectations as each level.

To conclude the study, a semi-structured post-scoring interview was conducted face-to-face and captured the beliefs and opinions of the teacher as it relates to the scoring and instruction of writing utilizing. The semi-structured interview protocol was audio-recorded for transcription and analysis to explore the influences of teachers' scoring decisions.

Data Analysis

Initially, it was intended that categories for coding would be derived from the research questions, the purpose of the study, and data collected and analyzed, but in further research and study, I came across the research study of Wolfe and Feltovich (1994) who considered the cognition of the scorer during the evaluation process. In their study, Wolfe and Feltovich collected data from novice and expert scorers as the participants scored student writing for a large-scale standardized assessment using a 6-point holistic rubric. The research by Wolfe and Feltovich on interpretive frameworks, also known as cognitive representational structures, is an expanded model of scorer cognition based upon the information–processing model of scorer cognition Freedman and Calfee published in 1983, who considered the process by which scorers processed information as they evaluate writing. Using this framework as a starting point, I was able

to expand upon this work to further as I considered the influences of teacher's decision making when scoring. In order to better situate my findings, I first explain the key features of the Wolfe and Feltovice (1994) model and then describe how that informed my own study.

Wolfe and Feltovich (1994) present a model of scorer cognition, as it relates to individuals scoring student writing samples using a holistic scoring rubric. Their work shows how expectations for the assignment on the part of the scorer (Model of Performance) along with prior knowledge of the scorer (sources of knowledge) play a role in the processing actions (model of scoring) the scorer utilizes to make scoring decisions. As depicted in Figure 5, the model of scoring "is a conceptual mapping/information processing model of an essay scorer's decision-making process" (Wolfe and Feltovich, 1994, p. 31). According to their model of scoring, an individual creates an image of the written text as the go about the interpretation phase, and those components are then mapped to the scorer's personal model of performance in the evaluation phase. It is in the justification and ultimately documentation stage of the model where the scorer monitors the evaluative decision for fairness and accuracy. It should be noted that according to their model the scorer draws upon prior knowledge and experience at each step in the process of evaluating an essay.
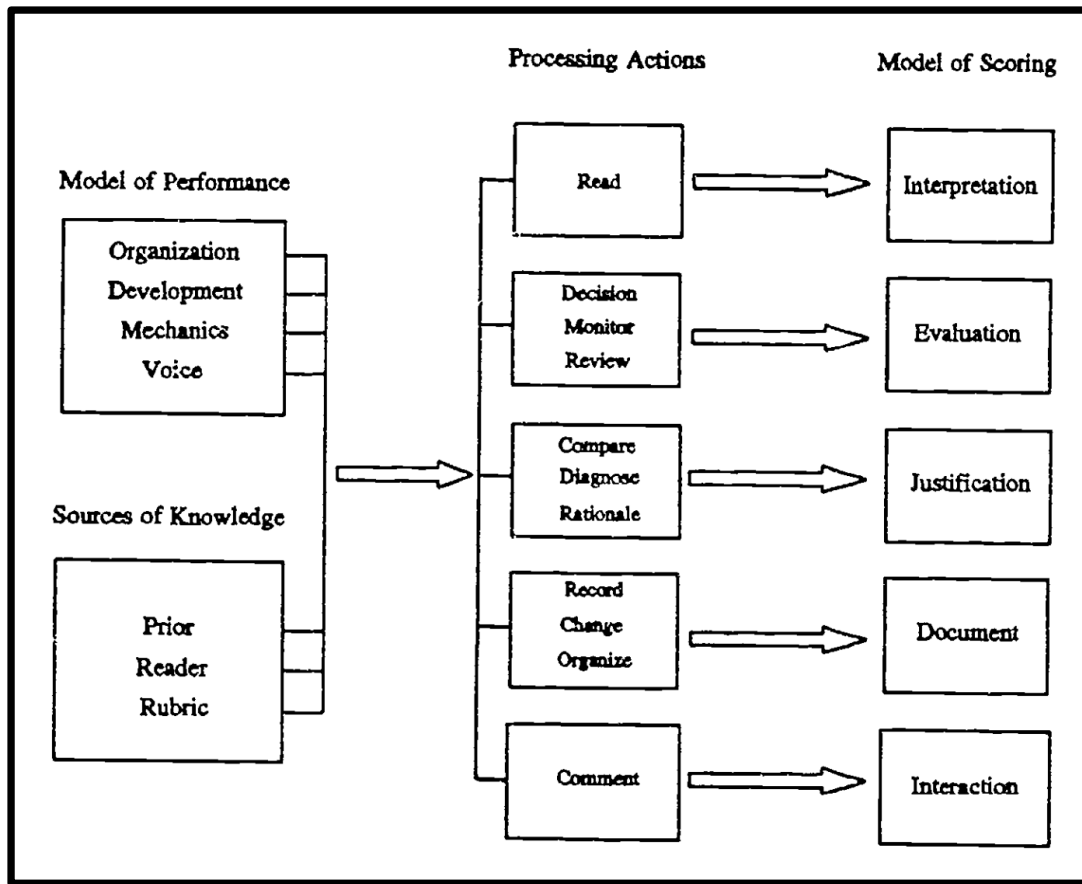
Figure 5. Wolfe & Feltovich (1994) Expanded Model of Scorer Cognition.

Models of performance. According to Wolfe and Feltovich, a model of

performance is "a cognitive representation of what constitutes proficient and non-

proficient" (1994, p. 14). For a writing assessment, a rater may have preconceived ideas

or expectations regarding structure, syntax, grammar, etc. And while these ideas or

expectations ultimately influence the evaluative decisions made by the rater it is unclear

how these ideas may influence the decision when teachers are scoring with and without a

rubric.

In this study, I engaged six teachers in two think-aloud protocol experiences as they evaluated student writing samples. The raters were asked to read and verbalize and thoughts they had while scoring each paper. Typically, the rater would read the piece aloud stopping along the way to make a comment about something that they read. In the first think-aloud teachers were not provided a rubric, this provided an opportunity for teachers to verbalize the thoughts and ideas for which they draw upon to make scoring decisions. In the second think-aloud teachers were given a rubric to use for the purposes of scoring. This time, although they still verbalized their thoughts and ideas, often those comments could be traced back to the language of the rubric. In both of the experiences, the comments made by teachers supported the concept of models of performance as Wolfe and Feltovich (1994) addressed in their research.

Because there was a specific rubric utilized for the second think-aloud-protocol, instead of utilizing the models of performance Wolfe and Feltovich (1994) utilized in their research, I choose to utilize the ones from the rubric the participants were using. Generally comments made by the educators could be placed in one of the four domains from The Texas Writing Pilot Rubric: 1) Organization (the overall focus and structure of the piece); 2) Content (support and elaboration); Language (relationship among ideas and use of literary style); and 4) Conventions (editing and sentence boundaries). While these four domains collectively accounted for the majority of the comments made by the teachers, there were some comments that fell outside these codes. As a result, the additional codes of: 1) Appearance (the way the writing looks); 2) Non-specific (general

102

comments about writing); and 3) Subject (compliance with the prompt or writing task) were also used by Wolfe and Feltovich (1994) in their study for the same reason were used.

Model of scoring. The second interpretive framework Wolfe and Feltovich (1994) utilized was the model of scoring, "a cognitive representation of the process through which one identifies and interprets evidence from a response and derives a score based on this information" (p. 16). This model demonstrates the manner in which a rater uses and manipulates the knowledge acquired from reading a piece, within the decision-making process, to arrive at a decision or score. The four main stages identified by Wolfe and Feltovich and used in this research study are: 1) Interpretation (actions used to make meaning of the text or to create a text image); 2) Evaluation (actions used to align and compare the model of performance and the text image); 3) Justification (actions used to provide a rationale or confirm the accuracy of a decision); and 4) Interactive (additional or peripheral information regarding the scoring experience).

Processing action. Wolfe and Feltovich (1994) initially believed that the model of scoring would be executed like a script by the scorers, but in their pilot study, they realized each scorer executed the process, or script, differently. For this reason, they identified processing actions that are associated with the model of scoring (see Table 4) to be used in combination with the model of scoring as a way of completing the scoring task. "A processing action is one of several cognitive activities that a scorer may perform when making a scoring judgment" (Wolfe & Feltovich, 1994, p. 17). Through their

research, Wolfe and Feltovich were able to identify a number of cognitive tasks

associated with each model of scoring that was also the basis for this study.

Table 4

Processing Actions as they relate to the Model of Scoring

Model of Scoring	Processing Action
Interpretive Actions having to do with obtaining information.	• Read (creating a text image)
Evaluative Actions having to do with a decision.	• Decision (assign a score) • Monitor (explain how the text image aligns or compares to the model of performance - during reading) • Review (survey how the text image aligns or compares to the model of performance - after reading)
Justification Actions having to do with providing a rationale for a decision.	• Compare (compare elements of the text to other sources of information) • Diagnose (provide ways the piece of writing could be improved) • Rationale (explain how the piece of writing exemplifies the model of performance)
Interactive Actions having to do with personal insights about the rating and reading task.	• Comment (personal comments regarding the process, text, or the writer)

Sources of knowledge. With the main question of this study being 'What are the

various factors that come into play as a teacher makes scoring decisions when evaluating

student writing?' I particularly wanted to consider the sources of knowledge educators

draw upon as they are making evaluative scoring decisions. Although Wolfe and

Feltovich (1994) identified sources of knowledge to be a consideration within their

expanded model of scorer cognition, I wanted to more thoroughly explore this area. In

their model, Wolfe and Felotvich identified three mediums from which the sources of

knowledge derive: 1) Prior (other papers previously read by the scorer); 2) Scorer (other

scores assigned by other scorers); and 3) Rubric (descriptions provided in a rubric). For

the purposes of this research study, I wanted to consider all sources of knowledge a

teacher may draw upon. For this part of the model I altered the definitions of these three

categories to allow new sub-categories to emerge. These sub-categories will be addressed

in the findings.

For the purposes of coding for this study, I coded each individual thought unit

from the think-aloud protocols (Ericsson & Simon, 1984) using the categories established

by Wolfe and Feltovich (1994) within their models of performance, models of scoring,

and processing actions. The data was then triangulated using the participant's self-

reporting during the initial and follow up interviews using Wolfe and Feltovich's sources

of knowledge as a starting place. A table embedded in the report of findings is denoted in

the results. The use of coding provided guidance in understanding and making meaning

of the data collected as a result of a thorough analysis of the information (Erlandson et

al., 1993). Further analysis included a review of patterns found within the codes to

identify any additional sub-categories and overarching themes.

Trustworthiness

Bracketing requires a researcher to put their experiences aside, to the greatest extent possible in order to take a fresh look at what is being studied and maintain reliability for the study (Crotty, 1998). While it should be noted that I do not supervise any of the teachers who participated in the study, as an initial creator of the design of the state pilot and as an active participant with The Texas Writing Pilot as a liaison at the educational service center level, the researcher acknowledge the need to bracket my experiences as they relate to the implementation of the statewide pilot.

Findings

By using a think-aloud protocol (Ericsson & Simon, 1984), participants were asked to verbalize the process and any thoughts they had while reading a student writing sample and determining a score for that essay. The audio of participants was digitally recorded and later transcribed for analysis. Transcribed protocols were coded in a similar fashion to the method outlined in the research of Wolfe and Feltovich (1994).

From the data analysis, I concluded that teachers draw from a wide range of influences and strategies when making scoring decisions. While the expanded model of scorer cognition developed by Wolfe and Feltovich (1994) proved to be a useful tool when considering the various factors that come into play as a teacher makes scoring decisions when evaluating student writing there is opportunity to further expand the model. This study supports the cognitive representation models Wolfe and Feltovich developed for models of performance and models of scoring; however, the data also

provided additional insight into an expanded view of source knowledge and processing actions.

For the purpose of this study, the data collected regarding influences for how teachers approach scoring student writing were mapped to models of performance as well as source knowledge. Data collected regarding the metacognitive strategies teachers use when making scoring decisions was mapped to processing actions. The remainder of this section will further elaborate upon the findings from the lens of each cognitive representation model.

Model of Performance

The model of performance is what the scorer constitutes as proficient or non-proficient performance. For a writing performance assessment, the model of performance may be the rules or expectations a scorer has for a particular piece or style of writing that influences the scoring decision. For example, when one of the participants said, "if I was grading this on persuasion it would probably get a low score because it's not very persuasive," she was taking into consideration her expectations for how the structure and organization of the piece contributed to the style of writing the student was composing – persuasive. For the purposes of this study, every individual thought, such as the one just shared, made by a rater during the think-aloud without the analytic rubric and the think-aloud with the analytic rubric was first coded to the four domains from The Texas Writing Pilot Rubric. In the case of the previous example shared, it was coded as structure, within the organization category. Although the first think-aloud protocol did

107

not use a specific rubric, the content that the participants drew upon for their model of performance could be easily identified within one of the four main expectations for writing within the rubric (organization, content, language, and conventions). Table 5 delineates each of the sub-categories within each of the domains of the analytic writing rubric. When instances arose where an individual thought unit was not able to be coded to one of the four main domains, (e.g., "here's one that is typed") the category of other was used and then later divided into additional sub-categories.

Table 5

Model of Performance

Organization	Content	Language	Conventions	Other
• Overall Organization • Structure • Focus • Progression • Central Idea	• Overall Content • Support • Elaboration	• Overall Language • Diction • Literary Devices • Varied Sentence Structure • Relationship Among Ideas	• Overall Conventions • Conventions • Editing • Sentence Boundaries • Paragraph Breaks	• Appearance • Subject • Non-Specific • Rubric

Model of Scoring and Processing Actions

In a second pass of the data, every individual thought made by a scorer during the think-aloud without the analytic rubric and the think-aloud with the analytic rubric were then coded as a processing action according to the action being executed by the scorer.

Processing actions are the ways in which the scorer manipulates a piece of knowledge during the scoring process. From there, each processing action was mapped back to one of the scoring elements within the model of scoring.

As I began coding processing actions, I closely utilized the eight processing actions identified by Wolfe and Feltovich (1994). In doing so, however, I began to discover that there were nuances within processing actions that needed to be teased out further. Table 6 denotes each of the processing actions that were used as codes for this study. The table also shows the relationship between processing actions and each scoring element that was identified within the model of scoring. The processing actions noted with an asterisk were not initially identified by Wolfe and Feltovich (1994).

Table 6

Expanded Processing Actions related to Model of Scoring Elements

Model of Scoring Elements	Processing Actions
Interpretive Actions having to do with obtaining information.	• Read • Question the text* • Respond to the text* • Respond to author's craft* • Compare to other sources of knowledge*
Evaluative Actions having to do with a decision.	• Decision • Monitor • Review • Compare to other sources of knowledge*

109

	• Diagnose
Justification Actions having to do with providing a rationale for a decision.	• Compare to other sources of knowledge* • Diagnose • Rationale • Recommend
Interactive Actions having to do with personal insights about the rating and reading task.	• Comment (general in nature) • Comment to the student* • Comment to the rubric*

Interpretive. Under the interpretive classification of processing actions, Wolfe and Feltovich (1994) defined 'Read' as the actual act of reading the text to create a text image. From the data, I found that as teachers engaged with the text, their process for interpreting the text was more nuanced than merely decoding words on a page to create an image of the text. In the process of creating a text image, there were moments where the rater would ask a question to make meaning of what they were reading, as well as times where the rater would comment or respond to the text directly as if to have a conversation with the text itself. For example, one rater while reading stopped to ask aloud, "is this a persuasive essay?" The significance is not that the rater was asking someone specific in the room, but rather asking as a way of making meaning fo the text. Another way the teacher would connect with the text was by commenting on how the author composed the piece of text, "I can see you worked really hard on description, and that is good." It is interesting that although the student herself was not present in the room, the rater made comments directly to the student as if she was present. Finally, it was not until a third pass with the data, and debriefs with peers, that I discovered

educators were also comparing and connecting to other sources of knowledge while creating their text image within the interpretive phase. One of the teachers situated the paper he was reading with grade level expectations when he remarked, "alright, I'm seeing high language use for a seventh grader." Ultimately, I identified four additional actions to better show the specific actions the participants took as raters interpreted each essay. Table 7 shows the processing actions for the interpretive element within the model of scoring as well as corresponding examples.

Table 7

Examples of Interpretive Processing Actions

Interpretive	
Additional Actions	Examples
Read Read the text.	"Teacher re-reads last sentence, paragraph 1 of Student Essay 1."
Question Ask a question in order to make meaning of the text.	"I'm not sure what she is saying, is she making a recommendation?"
Responding to Text Make a comment either to or about the text in order to make meaning of the text.	"Its stives to do better not strides to do better."
Responding to Author's Craft – Make a comment to or about the author in order to make meaning of the text.	"I love the use of onamonapia."

Comparing to Other Sources of Knowledge – Make a connection to other sources of knowledge in order to make meaning of the text.	"So, it Sounds like I'm listening to my wife a little bit because she loves animals."

Evaluative. According to the model presented by Wolfe and Feltovich (1994), once a rater creates an image of the text they are then able to move into the process of evaluation. As the raters go through the evaluation process, actions such as reviewing the text, making a decision, and monitoring that decision, are processing actions a rater uses when evaluating a piece of writing. The data from this study showed that while teachers did rely on the processing actions of decision, monitor, and review, that Wolfe and Feltovich (1994) identified, there were other actions a teacher would draw upon while making an evaluative decision.

At times, teachers would make a comparison about the paper they read to another form of writing, say another type of assessment. Initially, the processing action of compare only resided within the justification element of the model of scoring, but I came back each time to ask myself, is this statement a justification or not, I ultimately decided compare was also an evaluative action.

Teachers also made comments of diagnosis as part of their decision-making process. These comments would not fall in the justification element because they were not about why the piece got the score they decided upon but rather these comments were part of the decision-making process that ultimately contributed to the rater's decision.

Table 8 shows the processing actions for the evaluative element within the model of

scoring as well as corresponding examples.

Table 8

Examples of Evaluative Processing Actions

Evaluative	
Processing Actions	Examples
Decision – Assign a score.	"I would probably give it a two or three of the limited our basic in that rather than very limited because you can tell there are paragraphs."
Monitor – Explain how the text image aligns or compares to the model of performance - during reading.	"There's a lot of broad generalizations here."
Review Survey how the text image aligns or compares to the model of performance – after reading.	"It's very lofty ideas without anything specific, so I would say that the details and examples aren't relevant really so too vague."
Compare Relating the scoring decision to other sources of knowledge.	"I'm just think of it in terms of TELPAS like how low it would be this would be."
Diagnose – A determination of what could be done for improvement.	"It definitely needs more support possibly some some specific examples the details aren't really detailed."

Justification. Once a teacher would evaluate a student writing sample and

provide some kind of definitive score as part of the evaluation phase, there would be a

moment or two where the teacher would look back over the paper to confirm or revise

their scoring decision. These thought actions were coded with a processing action of

compare, diagnose, rationale, or recommend. Comparisons would happen when the scorer would draw upon other sources of knowledge and compare the paper they were reading to something else as a means of justifying the score. When a teacher would make a diagnostic statement they would often identify a portion of the paper and explain the shortcoming or how it could be improved. The data also revealed a level of uncertainty from the educators when making scoring decisions, so often the teacher would go back through the piece and point to a line or phrase that would justify or provide a rationale for why they chose to score it the way they did.

Finally, the processing action of 'Recommend' is another action that was an addition to the original classifications from Wolfe and Feltovich (1994). While the diagnose classification, states what needs to be improved, as a way of justifying the evaluation, recommend, goes a step farther to position the evaluator as a mentor to the writer and provides suggestions on not just what needs to be improved, but how to go about improving it. This is an important aspect to consider for it sheds some light on how the educator rater sees herself with pluralistic responsibilities. Table 9 shows the processing actions for the evaluative element within the model of scoring as well as corresponding examples.

Table 9

Examples of Justification Processing Actions

Justification	
Processing Actions	Examples
Compare – Draw upon other sources of knowledge as a means to justify the score given.	"And that's really a hard thing for kids at this age."
Diagnose – Identify a portion of the paper and explain the shortcoming or how it could be improved.	"There are not good transitions there between paragraphs."
Rationale – Provide a justification or explanation for why the score was given.	"They are not strong sentences but they do what they need to do."
Recommend – A suggestion for what needs to be improved and how it could be improved.	"Probably could have thrown in some other kinds of examples, you know political or something from other walks of life,"

Interactive. As scorers would read the student writing and offer evaluative judgments about a given piece, participants would make additional comments that did not did not directly map to the other three scoring elements. These comments, reflected in Table 10, vary in significance from statements like "hmmm," when participants had stopped to consider an aspect of the evaluation process or other comments where the scorer diverted from the evaluation process to make a connection to their own life. In the initial model, 'comment' was the only processing action within the Interactive scoring

element. The statements within the comment processing action generally provided information about any number of aspects regarding the rating experience. Upon further inspection, the data from this study revealed not all the comments made were directed to an emotionless audio recorder. Some of the comments that were made were directed to someone specifically, as if that author was actually in the room, while other comments were about or directed to the rubric itself. In reflecting on a writing sample, she had just read, on rater identified that the student had a basic introduction. She continued on in her reflection by explaining what she would do to help the student improve the introduction. "I have different ways to teach [introductions]. We have a badabing thing and I would probably show you badabing as a way to start that off the voice …" It was evident from the comments the teacher raters made directly to students who were not even present that the evaluator does not separate or discontinue her role as an educator just because she is scoring papers. While not an area of focus for this study, comments to the rubric may provide a glimpse into the rater's processing actions of the rubric itself.

Table 10

Examples of Interpretive Processing Actions

Interpretive	
Additional Actions	Examples
Comment – Provide information about any number of aspects regarding the rating experience.	"Hmmm" "and then we have one that is typed" "When we were writing our district rubric We had a lot of debate about exactly that what does that mean"

Comment to Student – Information about any number of aspects regarding the rating experience directed to the author of the writing.	"...but you worked really hard it a description which is good I can totally see you getting out of the pool and what you look like."
Comment to Rubric – Information about any number of aspects regarding the rating experience directed towards the rubric itself.	"Okay, so it's the same same thing you're looking for in that section for accomplished an exceptional." "For Conventions, I can only choose two four six I want to give her three but that's not an option."

Sources of Knowledge

Wolfe and Feltovich (1994) identified sources of knowledge as: 1) Prior - other papers the rater had previously scored; 2) Scorer - scores that might have been assigned to other papers; and 3) Rubric - descriptions provided by the rubric. Through my analysis, I was to determine what sources of knowledge do not solely reside within the evaluation experience as Wolfe and Feltovich (1994) identified. Sources of knowledge come from a variety of places, including the evaluator's personal experience, what they know about certain groups of students, as well as their knowledge of writing tasks, other assessments, and scoring rubrics.

In order to gain insight on the sources of knowledge that an educator draws upon when evaluating student writing, I reviewed each individual thought unit from the data and looked for instances where the rater made a connection to something outside of the essays or The Texas Writing Pilot Rubric. From there I looked for trends where I could subdivide the Sources of Knowledge Wolf and Feltovich (1994) identified into other

117

distinguishable categories. Table 11 shows the additional sub-categories identified from the data. For example, when one of the participants said, "You know what this feels like is an ESL kid." She was referring to the prior experience she has had teaching students acquiring English as their second language. This thought unit was coded in the prior category under student ability/language students. When another scorer stated, "That's definitely a pet peeve of mine that I noticed especially as a journal of journalism major." He was accessing his prior knowledge of personal experiences as a journalism major, so this thought unit was recorded as personal experience within the prior category. Ultimately, comments that accessed a scorer's knowledge from prior experiences included personal comments, comments about a student's ability, and instructional comments from experience teaching.

For comments coded in the scorer category, these comments demonstrated that the scorer was drawing knowledge from other types of assessments as well as other papers they had scored within the sample set used for this study. Sources of knowledge were also drawn from experience with other rubrics, as well as the rubric they were using for scoring within the study. Finally, an additional category was created called 'Assignment.' In reviewing the data it was discovered that teachers would make comments that would reference what they perceived to be is familiarity or an understanding of what the assignment was for the writing they were grading. For the purposes of this study, the directions that had been given to students for completing the assignment were not provided to the educators, but as they would read the selections they

would make assumptions about the paper based on what they thought they knew about the assignment requirements.

Table 11

Sub-Categories Identified for Source Knowledge

Source Knowledge	
Categories	Sub-Categories
Prior	• Personal Experience • Student Ability / Age • Student Ability / Language • Student Ability / Special Education • Instruction / Writing Technique • Instruction / Personal Expectations for Writing
Assessment	• Assessment Directions
Scorer	• Relationship to Other Papers • Relationship to Other Assessments
Rubric	• Rubric Expectations • Other Rubrics

Discussion

This study explored the scoring decisions teachers make when evaluating student writing and considered the various factors at play when these decisions are made. Using qualitative methods (Merriam, 2009), information was gathered to provide insight into how teachers assess student writing specifically for the purpose of considering the teacher's participatory role in assessing statewide writing assessments. Analysis of verbal statements recorded utilizing a think-aloud protocol as teachers scored student writing with and without a rubric provided a means for which to consider the decisions teachers

119

made as they scored. The expanded model of scoring cognition by Wolfe and Feltovich (1994) was used to code and make meaning of the data.

Two things to consider about the expanded model of scorer cognition developed by Wolf and Feltovich (1994) is that this model was based from research conducted pre-NCLB as well as with a holistic scoring rubric. The data from this study shows that even in an assessment centric culture the model, although slightly adapted, still provides relevance for considering the decision-making and processing actions of teachers while scoring student writing. Secondly, the data also supports that this model is applicable to an analytic rubric as well as a holistic rubric.

By examining the decisions teachers made when scoring student writing, and considering the approach they took to evaluate the writing, new light was shed on how various factors such as teachers' prior knowledge and expectations of writing come into play as teachers are making scoring decisions, as well as the metacognitive strategies teachers use when evaluating student writing. It was through the analysis of the metacognitive strategies teachers use that the data revealed teachers have a connection to the writer through the writing, even when evaluating writing from students who are not their own students. The data revealed ways in which teachers focused on instruction and possible future lessons for instruction based on the assessment. In each of the phases of scoring, the data revealed that teachers do not separate or suspend their role as an educator just because they are scoring an assessment. Instead, the teacher was able to

connect how the assessment plays into instruction by identifying next steps for the author to continue to grow as a writer.

Most notably, the findings confirm that the process of scoring student writing is a complex process of interpreting evaluating, reviewing and justifying one's score that draws upon a variety of factors and experiences. For example, one of those influential factors is Source Knowledge, or the prior knowledge of a teacher and the data revealed that the types of knowledge from which a teacher pulls from can vary from teacher to teacher. As the state continues to look to expand opportunities for alternative assessment types, such as The Texas Writing Pilot, more work is needed to be done in considering professional learning experiences that may influence or override some of what an educator brings to the scoring decisions so that interrater reliability might be improved.

Limitations and Delimitations

The findings of a qualitative research study (Merriam, 2009) are dependent upon interactions with participants (Creswell, 2013); therefore, some limitations could be formulated. For example, the data collected was dependent upon the participant's verbal participation in both the interviews and the think-aloud protocols, but through standard research protocols and clear directions for the participants, the effort was made to uphold the integrity of the study. Additionally, the data collected represents a few individual cases that may not be representative of the greater field of Texas writing teachers. However, it is my assertion that the findings have added value to the field of research as

they can transferable to other studies. There is still much room within the research on this topic further studies relating to teacher decision-making within performance assessment.

Summary and Conclusion

According to the interrater reliability study conducted by the state of Texas, "The correlations and rater-agreement of scoring [from The Texas Writing Pilot] never reached the same level as STAAR, at scale. While there were some sporadic highlights across the population in both Year 1 and Year 2, the overwhelming variance in data suggests that appropriately training enough educators to be standardized scorers would not be possible" (Texas Education Agency, 2018, p. 21). The problem with the study conducted by the TEA was that concluded there was no interrater reliability – case shut – never taking sufficient time to considered why this might have been the case. One thing to consider, which this study did, is the various factors that come into play as teachers make scoring decisions. By considering the various factors that come into play when teachers make scoring decisions as well as the insight gleaned from this study on the expanded model of Wofle's and Feltovich's scorer cognition model (1994), further research could potentially provide strategies to mitigate the scorer discrepancies by providing professional learning and training for educators that would assist in scorer calibration.

Influences of Teacher's Personal Experiences When Scoring Student Writing,

JoLLE@UGA

Working with teachers in the field of writing instruction and assessment is a familiar place for me. But on one particular fall day back in 2017, when I found myself standing in front of a room full of educators ready to hear from me how they would be scoring a new type of writing assessment, it did not quite feel like that familiar place to which I have grown accustomed. While I knew that although teachers were all physically attending the same training I was facilitating about assessment scoring procedures, I quickly came to realize all of the teachers sitting in the room had different experiences, expertise, and knowledge when it came to writing assessment and even writing instruction for that matter.

These teachers, who were participating professional development, were there because their district had agreed for their campus to participate in The Texas Writing Pilot, an alternative writing assessment pilot organized by the TEA. The goal of this pilot was to explore an alternative assessment with the goal of possibly replacing the state's current high-stakes writing assessment. In my role with the project, I supervised the implementation of the pilot in the Dallas-Fort Worth area of the state. In this capacity, I served as a liaison between TEA and participating districts, as well as facilitated the assessment scoring training TEA created for the teachers.

The training I facilitated had been created and given to me to deliver by the Student Assessment Division at TEA. As I watched teachers interact throughout the

training, I felt as though the information presented in the training had been created on the premise that all the teachers had started the morning on the same page of a long novel, when in reality I knew that was not true. As I continued to work with the teachers, I wondered what filters each of them were using to funnel the information I was sharing as they worked to make sense of the new learning for themselves. Throughout the course of the TEA's pilot study as a whole, I found a number of questions whirling in my head about statewide implementations and quality assessments, but ultimately all of my questions led me to want to further explore how teachers score student writing and what influences their scoring decisions.

Background on the Pilot

With pressure from parents and educators across the state concerned with the amount of testing students were having to bear, the state of Texas passed legislation in 2015 requiring the TEA to conduct a study related to authentic writing assessments. With an intended outcome of this alternative assessment being an evaluation of a student's growth in writing over the course of a school year, a key feature of Texas Writing Pilot was to provide students with more timely feedback on their writing within the context of classroom instruction. Ideally, students would have access to this feedback so they might be able to use it as they continued to improve their writing skills throughout the year. The pilot made provisions for the students' classroom teacher to assess student writing and provide timely feedback. Unfortunately, through a feasibility study conducted by the TEA, which measured "the quality of locally-produced ratings" (Texas Education

124

Agency, 2017, p. 5), the TEA found the results of teachers' scoring was not consistent enough for this type of assessment to be used for the purposes of high-stakes assessment. This led me back to my questions about teacher experience and knowledge. I wondered if something could be done to mitigate the scoring differences so that this type of assessment could be used for the purpose of high-stakes assessment.

In the TEA study, "the quality of locally-produced ratings" (Texas Education Agency, 2017, p. 5), was solely judged by the establishment of inter-rater reliability across scorers, even though there is significant research demonstrating that variability of scores is only one factor for consideration when judging consistency in ratings (Huot et al., 2010; Jeong, 2015a; Yancey, 1999; Zhang, 2016). For example, some studies suggest that a mitigating factor for inter-rater reliability can be the training raters receive that can drastically influence reliability on the part of the rater (Jeong, 2015a; Knoch et al., 2007). Even when raters make an intentional effort to follow the rubric as written, studies show that invariably the rater's decision-making process can be influenced by the overall impression of the piece of writing, and even personal intuition (Jeong, 2015a). In order to better understand why teachers did not produce consistent scores factors such as a teacher's experience and background must also be considered (Jeong, 2015a).

Purpose

The purpose of this qualitative study (Merriam, 2009) was to consider the sources of knowledge teachers draw from when evaluating student writing. While other studies have looked at a teacher's ability to evaluate student writing, these studies often focus on

125

quantitative data and focus on inter-rater reliability. By considering the knowledge and

professional learning (Birgin & Baki, 2007) that contributes to the different approaches

individual teachers use when scoring student writing samples; further light may be shed

on features of professional learning that influence teachers' scoring decisions. The

research question informing the study was, "What sources of knowledge do teachers

draw on when evaluating student writing?" The literature review that follows considers

the influences of teacher decision-making as well as components of professional learning.

Literature Review

Every day teachers must take part in a delicate dance of negotiating one's own

beliefs, curricular constraints, and institutional constraints (Borko et al., 1981; McMillan,

2005). For this reason, any action made on the part of the teacher, either conscious or

unconscious is the result of a complex decision process (Shavelson, 1973). As teachers

go throughout the school day, they must carefully weigh the internal and external

competing forces in their decision-making process (Black & Wiliam, 1998; McMillan,

2005). With professional learning being one of the influences of teacher decision-making,

this literature review highlights the intersection between decision-making and

professional learning.

Teacher Decision-Making

What an educator personally believes about children, education, and learning can

directly influence their decisions and behaviors within the classroom (Gill & Hoffman,

2009; McMillan, 2005; Putnam & Borko, 1997). Beliefs can also develop by one's own

126

experiences with teaching and learning (Gill & Hoffman, 2009). Everything from instruction to classroom management can be heavily influenced by a teacher's personal beliefs (Putnam & Borko, 1997), and yet teachers often struggle to explain how their beliefs influence their actions or decisions (Gill & Hoffman, 2009).

Outside the four walls of his or her classroom there are other factors, in addition to their beliefs, that come into play as teachers make decisions. A teacher must also consider curricular and institutional constraints such as school or district policies, expectations for the curriculum, as well as other mandated initiatives (Clark & Peterson, 1984; Griffith, et al., 2013; McMillan, 2005). This does not even take into account the pressure teachers feel from state mandated tests for their students to perform at a certain level (McMillan, 2005).

Unlike instructional decisions that are often made in the moment, teachers typically have more time to make assessment decisions. It is within this time that they are able to draw from a number of ideas and influences as they engage in the decision-making process. (Borko et al., 1981). The following section expands further on teacher decision-making by exploring the ways teachers develop their knowledge as an teacher through professional learning.

Professional Learning

For decades professional development experts have cautioned against traditional delivery models of professional learning. Teachers need opportunities to engage in just-in-time, job-embedded professional learning (Guskey & Yoon, 2009). When professional

development is customized to the needs of teachers and sustained over time there is evidence that it can positively contribute to improved teacher practice (Borko, 2004). In order for professional learning to prove meaningful for teachers, developers of professional learning must consider conditions for success. Optimal conditions for the development of educator knowledge include: 1) opportunities to study by participating in experiences within the core discipline and reflecting upon that work; 2) collective inquiry with other teachers that focuses on improved classroom practice; and 3) engagement in professional discourse around theory-based research (Darling-Hammond, 2008; Marlink & Wahleithner, 2011; Stokes, 2010). Through meaningful professional learning, teachers can deeply and flexibly develop an understanding of content and in turn use their newly acquired learning to create meaningful learning experiences for their students (Darling-Hammond, 2008).

For any given grade and subject area, the complexity and sophistication of the knowledge and skills an educator must possess to teach is immense. College pre-service education preparation programs serve as a beginning for learning the craft of teaching, but it is just the beginning because the adage of practice makes perfect is not just a meaningless phrase. While pre-service programs can provide a solid foundation they cannot replace the experience one can get by being a teacher. Additionally, by providing teachers time during a professional learning experience to reflect on the learning in light of current district and campus initiatives, teachers are better prepared to implement the new learning (Penuel et al., 2007). When teachers are able to engage and actually

experience (Neuman & Cunningham, 2009) the learning first hand and then reflect upon it, they are able to have a deeper understanding of the core features of the structural components for the given content (Penuel et. al., 2007).

When teachers engage in a professional learning community (DuFour & DuFour, 2013), collaborating with other teachers and engaging in discourse about student work, they begin to feel empowered to deepen their understanding of the content and ultimately adapt their teaching (Darling-Hammond, 2008). As teachers take part in collective inquiry with other teachers that focuses on improved classroom practice they are able to develop shared understanding as they learn from one another (Wilson & Berne, 1999). An important idea related to collective inquiry is that learning is not something to be packaged and delivered to teachers but rather teachers should feel empowered to actively construct their own learning experiences (Wilson & Berne, 1999) as the community engages in a particular line of inquiry to improve their practice (Borko, 2004).

Although quite often teachers seem to prefer to skirt around the conversation of theory-based research, it is widely known that teachers enjoy talking about issues and subjects relevant to their work and their students (Wilson & Berne, 1999). Some research suggests that this professional discourse is something teachers often do not have an opportunity to develop (Wilson & Berne, 1999) and may emerge as teachers further engage in collaborative professional communities. As educators engage in collaborative communities, teacher efficacy is often also improved and beliefs about teacher effectiveness is also influenced (Ross & Bruce, 2007).

When professional learning is inclusive of opportunities to participate in experiences within the core discipline, collective inquiry, and professional discourse around theory-based research, teachers are equipped to develop as professionals.

The optimal conditions for the development of educator knowledge explored in this section of the literature review is not a foreign concept to other statewide writing assessments. From 1990–1999 Kentucky utilized the KIRIS, a statewide standardized test comprised of a local performance assessment, short performance tasks, and extended time performance tasks that included writing portfolios (Abbott, 2016; Tung, 2010). An influential factor in the success of this assessment model was that training, for many of the state's teachers, took place through an initiative sponsored by the National Writing Project. It is reported that the targeted and intensive professional learning from the National Writing Project significantly contributed to scoring that was more accurate (Tung, 2010). This professional learning provided teachers the opportunity to engage in the discipline, participate in collective inquiry with other teachers that focused on improved classroom practice, and better understand the theory and research behind the project. Over time, it was found that a byproduct of this intensive training was improved instructional practice (Gomez, 1999). Because professional learning is one of the influences of teacher decision-making, this literature review considered the influences of teacher decision-making as well as explored the role professional learning plays in teacher decision-making.

Methods

This qualitative study (Merriam, 2009), situated within a constructivist research paradigm (Guba & Lincoln, 1994), considered the sources of knowledge teachers draw on when evaluating student writing. Grounded in the research question (Stake, 1995), the study provided insight on the phenomenon of scoring student writing through the collection of multiple data types (Stake, 1995). The research question guiding this study was, "What sources of knowledge do teachers draw on when evaluating student writing?" In the following section, I offer a brief explanation of the setting in which this research took place.

Setting and Participants

In order to begin to consider what could possibly be done to mitigate the scoring differences the TEA identified in The Texas Writing Pilot, this study focused on participants who taught writing at a grade level for which writing is assessed by the state (i.e., Grade 4, Grade 7, and Grade 9). Through the use of purposeful sampling (Merriam, 1998), six participants within the Dallas-Fort Worth metroplex, one from each of the three grade levels (Grade 4, Grade 7, and Grade 9), were selected for the study. Of those six teachers, three had more than 5 years of teaching experience and three had fewer than 5 years of teaching experience. All of them had diverse personal and professional learning experiences that influenced their scoring decisions. For the purpose of this study, all names used are pseudonyms.

Methods of Data Collection

In order to better understand and consider the sources of knowledge teachers draw from when evaluating student writing I began by asking each participant to complete an open-ended participant survey. This survey collected initial information about each participant such as years of experience, teaching certifications and initial background about their experiences with professional learning. Building from this survey, I arranged a time to meet with each teacher face-to-face and engage with them in a semi-structured interview (Merriam, 1998). This interview allowed me to get a better picture of their experiences in teaching writing as well as their experiences with professional learning related to writing instruction and assessment.

During the time of the initial interview, I also asked teachers to rate several student writing samples using a think-aloud protocol (Ericsson & Simon, 1984; Smagorinsky, 1994). In this particular exercise, while I did not provide them with a rubric, I did ask that they vocalize their thoughts, explain what they notice about the writing, and provide an explanation for how they might score the essay if they were to give it a grade. My aim was to gain insight about what the participants thought about and considered when making their scoring decision without being encumbered or guided by a rubric.

Following the teachers scoring of the student writing without a rubric, I then gave the participants the scoring rubric that was used for The Texas Writing Pilot. Because all six of the participants had not previously participated in The Texas Writing Pilot, they

132

had not seen this rubric. I asked the participants to take some time to get familiar with the rubric, before giving them several more writing samples to score – this time using The Texas Writing Pilot rubric. As I observed the participants and heard their comments during the think-alouds, I began to see patterns emerge for the sources of knowledge teachers draw from when making scoring decisions.

After some initial analysis, I met with each participant again for a post-scoring interview. A transcription of the interviews and the think-aloud protocols were typed and validated through member checking (Creswell, 2013). In this final interview, I asked each participant to review the transcription of them scoring the student writing samples. As we reviewed the transcripts together, I would stop to ask them about where they learned a particular idea or concept that they mentioned while scoring. For example, one teacher while scoring remarked about the quality of a students' use of subordinate clauses. When asked how she learned about subordinate clauses she shared about a specific time when she attended a training by Jeff Anderson. It is important to note that because of my role as a liaison between TEA and the participating districts in North Texas, bracketing (Creswell, 2013) was necessary and maintained by ensuring I kept solely within the scope of the pre-established protocols. Finally, I explain the methods followed to ensure the trustworthiness (Merriam, 2002) of the study.

Data Analysis

Scoring writing is a complex decision-making activity that calls for the scorer to draw upon several interpretive frameworks to make a scoring decision (Wolfe &

Feltovich, 1994). In order to investigate the knowledge teachers draw upon when

evaluating student writing, the primary unit of analysis was teachers' verbal statements

that revealed their processes of pedagogical reasoning and action (Shulman, 1987). To

begin the analysis process, I coded the data from both of the think-alouds teachers

participated in through the lens an interpretive framework Wolfe and Feltovich (1994)

identified as Source Knowledge. Source Knowledge is a comparison processing action

"performed by manipulating some external form of knowledge" (p. 36). In their study,

Wolfe and Feltovich (1994) identified three mediums from which the sources of

knowledge derive: 1) Prior (other papers previously read by the scorer); 2) Scorer (other

scores assigned by other scorers); and 3) Rubric (descriptions provided in the rubric).

Based on my own prior knowledge and experience of teacher's scoring student

writing, and to capture a broader picture related to teachers professional learning

experiences, I wanted to more fully explore this idea of Source Knowledge. To do this, I

chose to broaden the definitions of the three mediums Wolfe and Feltovich (1994)

originally identified to include: 1) Prior (any prior knowledge considered in the scoring

process); 2) Scorer (knowledge from experience scoring writing); and 3) Rubric

(knowledge from experience using rubrics for scoring). In addition to the three original

categories, I added a new category of: 4) Assignment (knowledge from experience with

writing assignments). The use of coding provided guidance in understanding and making

meaning of the data collected as a result of a thorough analysis of information (Erlandson

et al., 1993). By expanding the definitions and adding a new medium, insight was gained

134

into the knowledge teachers use to make scoring decisions by allowing new sub-categories to emerge.

Using the transcribed responses from the think-aloud-protocols, I was able to identify 11 different sub-categories of Source of Knowledge according to the comments teachers made. Below, Table 12 shows the categories and sub-categories identified by coding the think-aloud protocols and includes an example of each sub-category from the data.

Table 12

Sources of Knowledge

Categories	Sub-categories	Example
Prior Knowledge	Teacher's Personal Experience	"That's definitely a pet peeve of mine that I noticed especially as a journalism major."
	Student Ability - related to the age of the student	"If this was from a ninth grade class, it would get a low grade."
	Student Ability - related to English Language Learners	"I think this is almost an EL student, like an English Language Learner."
	Student Ability - related to Special Education Learners	"...but I think that they're an EL or maybe SPED, which if that was the case my grading would be a little bit... I mean I would know the student so I could take their accommodations into consideration."

	Writing Instruction/Writing Technique	"So she's probably being taught how to do [that kind of] description."
	Personal expectations for writing or writing instruction	"What am I looking for like some suspense…"
Assessment Knowledge	Relationship or connections to the directions of the assessment	"I'm wondering if there was an amount of words they had to use?"
Scorer Knowledge	Relationship or connections to other student writing papers	"So the difference between this one and that second one is what I was talking about - how that one didn't fit together and the pieces of this writing fit together…"
	Relationship or connections to other assessments	"You know three paragraphs is fine obviously she's having to fit it into the twenty-six lines that you're given for the STAAR test."
Rubric Knowledge	Relationship or connections to the rubric expectations from the assignment's rubric	"This one right away I know is different from the other one this is probably not gonna get any sixes or fives."
	Relationship or connections to expectations of other rubrics	"Because our school and many schools have to do a hundred point grading scale so, take a seventy, which is at our school which is the edge of passing..."

As an additional layer of the data analysis, I coded the transcripts from the post-scoring interview, where teachers self-identified their Source Knowledge from statements

they made as they scored student writing, using the eleven sub-categories of Source

Knowledge I had previously identified. In doing this layer of coding, I was able to

confirm the eleven sub-categories that I had identified in my initial coding. I was also

able to further explore the Personal Experience sub-category and identify nine more

narrowly defined sources of personal experience identified by the teachers. Table 13

provides a list of those nine sources.

Table 13

Types of Personal Experience

Experiences with Writing	Professional Learning Experiences	Peers and Mentors	Other
• Experience as a student in college or high school	• Structured learning experiences	• Peers and Mentor relationships	• Unknown/ not sure of source
• On the job experience as a teacher or in another profession	• Self-study and reading	• Team calibration experiences	
	• District lead professional learning	• Team/Commit tee curriculum writing	

It was through breaking the data down into these more narrowly defined

categories that I was able to better see patterns emerge from within the larger category of

Prior Knowledge. Using a thorough analysis, I was able to see that when scoring student

writing teachers were influenced by their own personal experiences with writing,

professional learning experiences, as well as peers and mentors. In the following section,

I provide a narrative of each participant and further analysis of these three influences.

Findings

In what follows, I discuss each of the three themes that emerged from the data: 1)

experience with writing; 2) professional learning; and 3) mentorship in terms of how the

teachers participating in the study identified the sources of knowledge they utilized when

evaluating student writing.

Experiences with Writing

Whether the experience is as a writer or as a teacher of writing, teachers are able

to draw from their experience to make meaning and understanding of the process of

writing. For an experienced teacher, the reciprocal process of teaching and then reflecting

on their teaching (Neuman & Cunningham, 2009) provides opportunity to establish

beliefs and behaviors about instruction and assessment. Even a first year teacher is not

void of experiences with writing. Whether or not she considers herself a writer, from the

time she entered kindergarten she was exposed to thousands of hours in a classroom as a

student observing and experiencing her teacher's instruction. Lortie (1975) used the term

apprenticeship of observation to describe the acculturation to education one receives

before even entering a teacher education program. It is this experience of apprenticeship

that penetrates the beliefs and behaviors of an individual, even years later, in her own

classroom. This section further considers how one's own experiences with writing plays a role in making writing scoring decisions.

Debbie, an experienced ninth grade teacher. With more than 20 years of experience teaching, Debbie has taught sixth – 12th grade English and language arts including AP and dual credit college courses. Her current assignment is teaching ninth grade students at a public high school that is a special career-oriented academy in her district. When asked to describe herself as a writing teacher she said "I'm very particular about form and function, but I'm also eager to teach students the rules so that they can learn to break them. My motto is, 'If it doesn't say you can't, then you can.'" In her 20 years of teaching, she recalls having been exposed to a variety of professional learning experiences including Laying the Foundations, Thinking Maps, AP Summer Institutes and a number of district-lead professional development including the Jane Schafer method as well as Writing and Reading Across the Disciplines. Although she identifies that these experiences have been helpful in her career, she feels ultimately it is her experience in the classroom and as a writer herself that has been the best teacher to her. When asked how she learned to teach writing, she said it was through the reciprocal process of teaching and assessing student writing where she learned more about how to teach. More than anyone else teaching or training her how to teach writing she explained it "was more just once I saw what [the students] produced from what I taught them and then I was like well, that didn't work, we're gonna try it this a different way. So it took several years to figure it out [how to teach writing]."

Rene, a novice ninth grade teacher. Rene, a first-year educator, teaches ninth grade English at a comprehensive high school just outside Dallas, Texas. As a writing teacher, she believes that although writing is a creative process, too often in school it is about whether or not it fits into a particular box. In reflecting on the process of writing, Rene said, "I think for writing to be done well, people – anyone not just students – have to learn that to revise and edit is ongoing and continuous, and it is okay to make mistakes." When asked how she has learned to teach writing, she simply stated trial and error, much like Debbie, the experienced ninth grade teacher, did. As she considered her professional learning experiences, she was only able to share one brief time after school where she graded a few of her student's essays with other teachers. Outside of this experience, she could not recall any other learning opportunity, formal or otherwise, that she had attended where a focus on writing instruction or assessment had occurred. Despite a lack of professional experience, Rene did say she often draws for her personal experiences in high school as well as college, where she took mostly English courses.

Both Debbie and Rene consider themselves writers. In many of their responses to the comments they made during scoring using the think-aloud protocol, both teachers could identify a teacher they had previously had who taught them how to do something in their own writing as a student. For Debbie, Mrs. Johnson in eighth grade, while for Rene, it was her professor freshman year of college. Also, both teachers referenced the importance of trial and error as a learning tool for how to teach writing. Both teachers explained how, by participating in experiences within the core discipline and then

140

engaging in reflective practice of their teaching, they are able to consider how to improve a lesson to ensure student mastery (Darling-Hammond, 2008; Neuman & Cunningham, 2009). In analyzing the teachers' comments, while both teachers talked about their experience as a writer and as a teacher, Debbie more often referenced her experience as a teacher whereas Rene referenced her experience as a student in school more often.

Professional Learning

It is no secret that teachers often have a less than enthusiastic outlook when it comes to professional development. But it is not because teachers inherently do not like to learn. In order for professional learning to be both meaningful and relevant for teachers, the teachers who participate, the facilitator who guides the learning, and the situation or context for which the professional learning is given must be considered (Borko, 2004). Ultimately, teachers find learning meaningful when they are able to collaborate with other teachers and collectively engage in efforts to learn and improve their practice. Through these types of experiences, they are able to develop shared meaning and understanding from one another (Wilson & Berne, 1999). These Communities of Practice, where professionals come together with a shared passion or concern and take steps to learn and share with each other (Wenger, 1999), become a powerful learning experiences for educators looking to improve their craft. This section explores learning experiences that teachers drew from when making writing scoring decisions.

Donna, an experienced seventh grade teacher. Although for the last 24 years Donna has taught middle school students reading and writing in the suburbs of Dallas– Fort Worth, her first career was actually a clerical position in the engineering field. It was in this position she was able to see the importance of quality writing. In her initial interview, she shared a memory of one particular engineer who was such a poor writer that he got marked down on his performance reviews, which ultimately cost him raises and bonuses. From her experience, Donna believes writing is "a fundamental skill that everyone should learn so that they can be a good communicator through written words," and it is evident that she approaches instruction from this stance as well. As I asked her questions in the initial interview about writing or writing instruction, Donna would often respond through the lens of both reading and writing. To her, she does not see them as separate disciplines, rather it is by becoming a better reader that directly helps you become a better writer. When reflecting upon what has influenced her teaching, she recalls attending the New Jersey Writing Project as the most influential learning experience she has ever attended. Although she attended during her first year of teaching, she was able to provide many examples of ideas and strategies she uses to this day. Initially, as Donna shared, she said that she really has not attended very much professional learning but as I listened to her talk, it was very evident that Donna is an avid learner and had engaged in a number of learning experiences – although not in the traditional sense of the term. While several times she mentioned personal study and research as one of her ways of learning new teaching techniques, one main source of her

learning has been her involvement in Jim Burke's English Companion Ning. It is on this social networking site where she has learned from others who share their experiences and recommendations on the site as well as by participating with educators from all over the world through a number of online book studies hosted by the site. For Donna, her experiences suggest that she finds real time learning more authentic than a formalized professional learning event.

Anthony, a novice seventh grade teacher. Anthony, a seventh grade English language arts teacher in the Dallas–Fort Worth suburbs, has 5 years of teaching experience. In college, he majored in journalism and spent some time in that career field before getting into education. Like Donna, Anthony identified the interdependent nature of reading and writing as one of his core beliefs, "you have to be able to read in order to write well and you have to be able to write in order to read well." Specifically when speaking about his middle school students he has found that teaching writing is most relevant to the students when they have opportunities to write about themselves. When asked how he learned to teach writing he acknowledged that he learned through other individuals who taught and modeled writing instruction for him. For example, in his first year of teaching he attended a 3-day institute hosted by his district that was modeled after the work of the National Writing Project. In reflecting on what was most beneficial about this experience he said, "I mean it was teachers teaching teachers and so they did a good job of showing us 'here's what to do,' 'here's why its important,' and 'here's why it works' so I really came away with a lot I will never forget." During our time together, he

showed me a notebook from this experience where he took all of his notes and still uses and references. In addition to the institute he referenced, Anthony also talked about attending district-led professional learning, AP Summer Institutes, and on-going learning and collaboration with his professional learning community on his campus.

It has been said before that on-going professional learning is more effective than a one-off professional development event (Guskey & Yoon, 2009), and both Donna and Anthony reinforce this idea. While Donna referenced the New Jersey Writing Project and Anthony referenced National Writing Project, it is the intention of both organizations to get teachers writing themselves as well as learning from other teachers. For Donna and Anthony, these extended professional learning experiences occurred early in their careers but are still something very much relevant in their teaching today. Also, while Anthony mentioned collaboration on his campus with his professional learning community (DuFour & DuFour, 2013) and Donna mentioned her participation on the English Companion Ning, both teachers referenced their involvement in communities of practice (Wenger, 1999), where they engaged with other teachers with the expressed interest of learning how to improve their practice as a teacher. This collective inquiry (DuFour & DuFour, 2013; Wilson & Berne, 1999) serves as an opportunity to engage in professional learning that is customized to the specific needs and interests of the individual or group (Borko, 2004).

Mentorship

Truly, teaching is a never-ending process of learning and improving; not only for students, but for teachers as well. While many learn from past experiences and professional learning experience, others find value and meaning when they engage with others to talk about issues and subjects relevant to their work and their learners (Wilson & Berne, 1999). This professional discourse can directly "influence the teacher's beliefs about their effectiveness" (Wilson & Berne, 2007, p. 54). As teachers feel supported by others they are able to have confidence in believing they could successfully implement the new strategy or skill. In this section, I explore how this idea of mentorship can take on both the form of a distant mentor (Hubbard & Power, 1993) as well as the teacher next.

Sarah, an experienced fourth grade teacher. Sarah is an experienced teacher of more than 20 years, teaches fourth grade reading and language arts in a school district just north of Fort Worth. As an experienced teacher in her district, Sarah is part of the district curriculum writing team and often delivers professional learning for her district as well. In her classroom she believes "it's super important for kids to feel like they're successful writers so more so than picking out a lot of the things that are their weaknesses, and really focusing on their strengths seems to help them get better." She wants her students to see themselves as writers and she learned the importance of this experience when she attended one of Lucy Calkins' workshops at Teacher's College in New York. In reflecting on her experience, she said that a big takeaway for her was that

145

"the whole time you are there you're a student, so everything you do is like your a kid in the classroom and it completely changed the way I approach my classroom." It was through this experience she realized how intimidating writing, and even sharing your writing can be for students. Because Sarah loves to read, she has read countless books on professional learning and even confessed that her cabinets are full of professional books about teaching writing. In talking about the professional books she has read, she referred to the authors more like mentors (John-Steiner, 1985) and explains that she gravitates towards the authors who she can relate to as a teacher because she feels like she can see how they do the strategy or technique in their own classroom. Regardless of where or how Sarah learns, for her it is all about finding ways that will ultimately make her students better writers.

Brittany, a novice fourth grade teacher. For the last 4 years, Brittany has taught fourth grade reading and language arts in a school district just north of Fort Worth. Even though she considers herself a writer, she feels like "it's one of those things that I like doing, but then sharing it is hard for me." However, when it comes to her students, she feels completely comfortable sharing her writing with them. Often, she will write in front of her students or with her students to model the kind of writing they are practicing. As she plans her writing instruction, she always tries to make writing "interesting and get [students] excited about it, so it's not something like this rote thing that they have to do every day." As we talked about how she learned to teach writing and various professional learning experiences she has attended, she was quick to name a strategy or idea and

recall, by name, the person or other teacher who had helped her learn that. One particular

teacher, who now teaches across the hall from her, came up a number of times as

someone she has learned a lot from specifically as it relates to writing instruction.

Brittany was able to identify numerous learning opportunities her district has provided for

her, but the ones she recalls being the most impactful learning experiences have been the

times where a more experienced teacher either has coached her or modeled instruction for

her. In her second year of teaching, she had the opportunity to participate in a writing

cadre which met several times throughout the year. During the times they met, other

teachers with more experience would come in to work with the cadre by showing or

demonstrating something that they were doing in their classroom. From there the

experienced teachers would also walk the cadre through student writing samples to show

how the strategy or technique they shared looked in student writing. Brittany shared that

this was meaningful to her because not only was she able to see a strategy in action, but

she was then able to go back to her classroom to practice it, before she met with the cadre

the next time to debrief the strategy.

A key component of professional learning identified in the literature review was

opportunities for educators to engage in professional discourse around theory-based

research (Ross & Bruce, 2007; Wilson & Berne, 1999). As I mentioned for Sarah, her

mentors come by way of the professional authors whose books she read. It was evident

that she consults these authors as expert educators and looks to them for sound research

based practice that she can use in her classroom. For Sarah, as she explained the close

personal relationship she feels to the author as she reads their book(s), she was also able to identify how this relationship stretched and deepen her understanding about the craft of teaching and assessing writing. John-Steiner termed this experience "the legacy of their distant teachers" (1985, p. 37) and explained that this distiant teacher, while can be important during the early years of one's experience, provides rich nourishment throughout one's entire career. Likewise, Brittany has also benefited from mentorship, although her mentorship came by way of peers in her district and on her campus. Whether it was the extended professional learning experiences with more experienced teachers came and shared their wisdom, or the teacher she sought advice from across the hall, it was clear that she seeks out these individuals so she can engage them on a professional level and in doing so she was able to see her own effectiveness as a teacher improve (Ross & Bruce, 2007).

Conclusions and Implications

By considering and identifying sources of knowledge teachers draw from when evaluating student writing, this study illustrates the complexity of teacher decision-making in the assessment decision-making process. As a result of data analysis, the data from six educators who participated in the study suggests that personal experience, professional learning, and mentorship are three components that influence teacher decision-making. This study provides a glimpse into the complexity of teacher decision-making. By examining experienced and novice educators from three different grade

levels, the study also was able to shed light on the impact experience in teaching may play in decision-making as well.

It was my experience through the interview process that educators could often adequately articulate what influenced the comments and decisions they made in the scoring process. As previous research suggests, the data from this study also supports the idea that teachers do not apply a rubric to a piece of writing in a vacuum without further consideration of external factors such as the student's ability, or even other assessment designs (Jeong, 2015a; Lumley, 2002). Ultimately, what educators believe about teaching and learning, can in fact, directly influence their decisions (Gill & Hoffman, 2009; McMillan, 2005; Putnam & Borko, 1997).

Teachers found a level of meaning and usefulness from the professional experiences that they mentioned, and because of that they felt empowered to turn and use the newly acquired knowledge in their classroom (Darling-Hammond, 2008). Although I used two teachers to highlight each finding of the data, all six of the educators to one degree or another referenced that their professional decisions were most influenced by experiences where they had opportunities to: 1) participate within the core discipline and reflecting upon that work; 2) engage in inquiry with other teachers that focuses on improved classroom practice; and 3) engage in professional discourse around theory-based research (Darling-Hammond, 2008; Marlink & Wahleithner, 2011; Stokes, 2010).

The teachers in this study, represent the grade levels where writing is currently assessed in the state of Texas as well as a mix of experienced and novice educators. Data

from the study demonstrates the various influences of teacher decision-making when it comes to writing assessment. Implications for a study like this include potential design considerations for professional learning experiences but these considerations must be made at the state while considering the impact at the teacher level.

From the data in this study, I was able to conclude that teachers most often draw from their experiences which include personal experience, professional learning and inquiry, and mentorship, or professional discourse. When scoring student writing, a teacher's experience is at the forefront of their decision making and the patterns discovered in this study reveal an important relationship between personal experience, professional learning and inquiry, and mentorship, or professional discourse. As a research community, further exploration is needed in how each of these elements work in concert with one another and ultimately influence teacher decision making.

For a state seeking to utilize teachers in the scoring of assessments for the purpose of high-stakes assessment, consideration to the timing and design of training must also include the knowledge and skills an individual teacher brings to the scoring table. As previous research suggests, a mitigating factor for inter-rater reliability can be the training raters receive because it can drastically influence reliability on the part of the rater (Jeong, 2015a; Knoch et al., 2007). While the findings from the study conducted by the TEA indicate that teachers did not score reliably consistent enough for this type of assessment to scaled statewide (Texas Education Agency, 2017), the reports also indicate that teachers, at most, were provided with a three-hour scoring calibration training (Texas

Education Agency, 2017). When teachers in this study reflected on learning that was most impactful in their decision-making process they often pointed to on-going, sustained professional learning, such as a week-long Advanced Placement Summer Institute, rather than a one day, one-off professional development.

Another type of learning experience that influenced teacher decision-making was when they were given time to practice within the content area and then reflect upon the learning. Additionally, research indicates that within the reflection process it is important to give time for teachers to consider how the new learning intersects with things they already know (Penuel et al., 2007). In the case of an experienced teacher who has had numerous experiences providing accommodated instruction or special considerations for a language learner or a student in special education, they must have the opportunity to mediate how something such as accommodations are applied or is not applied on a standardized assessment. Likewise, for a new educator, it cannot be assumed that seemingly subjective terminology found in a rubric such as "purposeful, logical, and highly effective transitions," "skillfully controlled sentences," or "control of sentence boundaries" have universally agreed upon definitions and these educators need the opportunity to reflect on these terms in context of the assessment itself.

Finally, during the scoring calibration training provided by the TEA, the rubric was introduced to teachers and they were asked to evaluate several sample papers. After teachers evaluated the papers the presenter would go over with the group and debrief the score the TEA had assigned each paper. This type of mechanic delivery of rubric

calibration does not align with educators who identified that they rely on the mentorship relationship to make meaning of their learning. While it does provide them with a model of what a score looks like, educators would also benefit from meaningful discourse about the research and theory behind the design and implementation of the scoring rubric. Furthermore, teachers would also benefit from the opportunity to converse with one another as a means of making sense of the rubric and scoring process.

This study illustrates the complexity of teacher decision-making in the assessment decision-making process and provides further consideration into the influences of scoring calibration such as teacher knowledge. This study also highlights the need for intentionally designed professional learning about scoring as a means to mitigate scoring differences and ultimately improve inter-rater reliability (Jeong, 2015a; Knoch et al., 2007). By identifying the various influences of teacher decision-making when scoring student writing, this study illuminates potential opportunities for the state in the design and implementation of the scoring calibration training.

CHAPTER V

DISCUSSION AND CONCLUSIONS

This chapter includes a summary of the qualitative study (Merriam, 2009) related

to influences of teacher knowledge and professional learning on teachers scoring student

writing assessments. Following a brief summary of the study, conclusions drawn from the

findings presented in Chapter 4 will be discussed. The final section of this chapter will

address implications for action and recommendations for further research.

Summary of the Study

As the state of Texas continues to seek ways to alternatively assess students,

specifically in the area of writing, there is a need for teachers to consistently and reliably

score student writing. Through exploring the various factors that come into play as a

teacher makes scoring decisions when evaluating student writing, themes emerged related

to influences of teacher's scoring decisions.

Overview of the Problem

Although states and school districts are living in an age of assessment and

accountability (Stecher, 2010), to meet federal and state accountability mandates, there is

still a need for authentic assessments that teachers can use for instructional purposes. In

order for the state to see alternative assessments such as The Texas Writing Pilot as a

viable alternative assessment to the current assessment and accountability systems,

writing teachers must be able to score student writing using a standardized rubric.

Although the determination of this feasibility typically rests on inter-rater reliability

across raters, there is much more to consider in the rating process (Huot et al., 2010; Jeong, 2015a; Yancey, 1999; Zhang, 2016). Thus, the problem of this study was to consider how a teacher's knowledge and training (Bírgín & Bakí, 2007) contribute to their scoring approach, through an analysis of the considerations and decisions they make when scoring student writing.

Purpose Statement and Research Questions

As the state explores alternatives to the current writing assessment and determines the scalability of an alternate assessment option for student writing, this study seeks to explore the various factors that come into play as a teacher makes scoring decisions when evaluating student writing. The purpose of this study is to explore the decision-making process when scoring student work samples and to consider the teacher knowledge and professional learning that contribute to the different scoring approaches teachers use. The following questions guided the qualitative study (Merriam, 2009):

1. What is the current state of the implementation of H.B. 1164 (2015) regarding performance assessment?

2. What are the various factors that come into play as a teacher makes scoring decisions when evaluating student writing?

   a. How do teachers approach the application of a rubric in evaluating student writing?

   b. What metacognitive strategies do teachers use in the process of making scoring decisions when evaluating student writing using a rubric?

Review of the Theoretical Framework

From the premise that we are a collective of our beliefs, experiences, relationships, the way others see us, and the way we see ourselves (Crotty, 1998) it would stand to reason that knowledge is a constructed endeavor, for it is the individual who possesses the power to construct truth or meaning from their interactions within the world (Crotty, 1998). John Piaget's work on constructivism explains that the acquisition of knowledge comes through an individual's interactions with the world, people, and things (Piaget, 2014). To build further on Piaget's work, Seymour Papert shed light on the influential role materials situated within a culture play in determining meaning-making (Papert, 1980; Papert, 2006; Picard, et al., 2004). While constructionism and constructivism share similarities, Papert (1980) believed that is that it is impossible for objects to fully be described apart from the person who is experiencing that object. because it is the experience the human has with the object that is the essential to what the object is and how it is used. Papert's theory of constructionism sheds light on how an individual access different media in various contexts to form and transform ideas in their mind (Harel & Papert, 1991; Papert, 2006). This idea that an individual can shape and build upon their ideas in context provides a poignant divergence to Piaget's stage theory (Ackermann, 2004; Papert, 2011).

Once teachers experience new learning of any kind, formal or informal, they have an opportunity to return to their classrooms and apply that concept or idea to their professional practice. Papert's interest in how a learner engages with various media to

155

boost self-directed learning and support in the development of new knowledge can be applied to teachers as they engage in the act of internalizing professional learning by way of enacting it out in their professional practice. For teachers, enacting or tinkering with an idea or concept learned, by practicing within the profession, to continue in the act of making knowledge for oneself improves one's competence in teaching (Duran & Tipping, 2017). This study investigated professional learning as a mediating tool for internalizing (Vygotsky, 1980) and transforming one's knowledge to understand better the decisions teachers make when evaluating student writing using a rubric.

Review of the Methodology

To explore efforts by the state of Texas to implement an alternative writing assessment as well as to examine the approaches individual teachers take when scoring writing, a qualitative study methodology (Merriam, 2009) was used. In order to situate the study and consider current efforts underway across the state to determine the teacher's role in assessing alternative writing assessments, a thorough document analysis (Gibson & Brown, 2009) was conducted of the current documents related to The Texas Writing Pilot published by the TEA. Following the document analysis, purposive sampling (Merriam, 1998) was used to select six Texas writing teachers, two from each of the following grades: Grade 4, Grade 7, and Grade 9. For each grade level represented, one of the teachers selected possessed more than five years of experience teaching, and other teacher selected for that grade level had fewer than five years of experience teaching.

156

As part of the study, each participant completed an online survey (see Appendix B) to obtain background information, an initial face-to-face interview, as well as a follow-up face-to-face interview. In addition to the interviews, teachers also participated in two think-aloud protocol experiences by evaluating student work without a rubric and then with a rubric while verbalizing his or her scoring decisions into an audio recorder. Through this think-aloud protocol (Ericsson & Simon, 1984; Smagorinsky, 1994), an effort was made to capture the cognitive process of the teachers' scoring decisions as well as the teacher's thoughts behind how the decisions are made (Zhang, 2016). Field notes were taken during interviews, and all interviews and think-aloud recordings were transcribed. Data analysis drew upon the thematic analysis method (Braun & Clarke, 2006). Findings and emerging themes were checked with the participants (Lincoln & Guba, 1985).

Major Findings

Following the collection and analysis of both the documents from The Texas Writing Pilot as well as the data from the teacher interviews and scoring sessions, the presentation of findings related to the influences of teacher knowledge and professional learning on teachers scoring student writing assessments were reported. These findings provide further insight into each of the research questions.

State of Implementation

The document analysis used to analyze the public documents pertaining to The Texas Writing Pilot shed light on the current state of the implementation of H.B. 1164

(2015) regarding performance assessment. Through an analysis of the data, the following

six categories were identified as important influences of the pilot: 1) Authentic

Assessment; 2) Writing Portfolios; 3) Student Growth; 4) Instruction; 5) Scoring: Rater

Agreement; and 6) Training. Findings and supporting evidence from each category,

presented in Article 1 – *Lessons Learned from The Texas Writing Pilot*, shed insight into

what happened during The Texas Writing Pilot. These findings revealed that the design

of the pilot was to be "a collaborative design process, inclusive of teachers, Education

Service Centers (ESCs), and institutes of higher education" (Texas Education Agency,

2016a, p. 2). It was the intention that this assessment design would "fully align the

assessment with local instruction and scope and sequence of curriculum" (Texas

Education Agency, 2018, p. 7) and assess student mastery and growth over the course of

a school year. To support teachers in the implementation of this assessment, training

would be facilitated from the TEA to local service center representatives. Through a

train-the-trainer model, the training would then be passed on to teachers through in-

person scoring sessions. Although teachers reported this type of assessment improved

classroom instruction, the TEA reported that The Texas Writing Pilot would not be a

viable assessment option because "scoring correlations and rater-agreement never

reached the same level as STAAR, at scale" (Texas Education Agency, 2018, p. 21).

Ultimately, the data revealed a clear disconnect between how educators teach writing and

how the state assesses writing. These dueling forces are ultimately what lead to the

breakdown between assessment, instruction, scoring, and calibration that I presented in the discussion.

Influences of Teacher Decision-Making

From the analysis of the interviews and think-aloud protocol transcripts in Article 2 – *Influences of Teachers' Scoring Decisions When Evaluating Student Writing*, I concluded that teachers draw from a wide range of influences and strategies when making scoring decisions. Using the expanded model of scorer cognition developed by Wolfe and Feltovich (1994), I was able to focus in on the various factors that come into play when teachers make scoring decisions as they evaluate student writing. In addition, the model provided insight into how teachers applied the use of a rubric to scoring and relied on metacognitive strategies when making scoring decisions. While the expanded model of scorer cognition (Wolfe & Feltovich, 1994) proved to be a useful tool when considering the various factors that come into play as a teacher makes scoring decisions when evaluating student writing, the data revealed an opportunity to expand the model further. The study supported the cognitive representation models Wolfe and Feltovich (1994) identified as models of performance and models of scoring; however, the data provided additional insight into an expanded view of source knowledge and processing actions.

Through the data collected, the model of performance, what the scorer constitutes as proficient or non-proficient performance, could be easily identified within one of the four main expectations for writing found within the rubric (organization, content,

159

language, and conventions). When instances arose where an individual thought unit was not able to be coded to one of the four main domains, the category of other was used and then later divided into additional sub-categories. In a second pass of the data, every individual thought made by a scorer during the think-aloud without the analytic rubric and the think-aloud with the analytic rubric were then coded as a Processing Action according to the action being executed by the scorer. processing actions are the ways in which the scorer manipulates a piece of knowledge during the scoring process. From there, each processing action was mapped back to one of the scoring elements within the model of scoring. In completing this analysis, I discovered that there were nuanced differences within processing actions than originally identified by Wolfe and Feltovich (1994).

Finally, to gain insight into the sources of knowledge that an educator draws upon when evaluating student writing, I reviewed each individual thought unit from the data and looked for instances where the scorer made a connection to something outside of the essays or The Texas Writing Pilot Rubric. Through my analysis, I was able to determine that sources of knowledge do not solely reside within the evaluation experience, as Wolfe and Feltovich (1994) identified. Rather, Source Knowledge comes from a variety of places, including the evaluator's personal experience, what they know about certain groups of students, as well as their knowledge of writing tasks, other assessments, and scoring rubrics.

160

Influences of Professional Learning

Findings from Article 3 – *Influences of Teachers' Professional Experiences When Scoring Student Writing,* shed light on the complexity of teacher-decision-making. This study highlighted the need for intentionally designed professional learning about scoring as a means to mitigate scoring differences and ultimately improve inter-rater reliability (Jeong, 2015a; Knoch et al., 2007). By identifying the various influences of teacher decision-making when scoring student writing, this study illuminates potential opportunities for the state in the design and implementation of the scoring calibration training.

As a result of data analysis, the data from six educators who participated in the study suggests that personal experience, professional learning, and mentorship are three components that influence teacher decision-making. By examining experienced and novice educators from three different grade-levels, the study also was able to shed light on the impact experience in teaching may play in decision-making as well. As previous research suggests, the data from this study also supports the idea that teachers do not apply a rubric to a piece of writing in a vacuum without further consideration of external factors such as the student's ability, or even other assessment designs (Jeong, 2015a; Lumley, 2002). Ultimately, what educators believe about teaching and learning can, in fact, directly influence their decisions (Gill & Hoffman, 2009; McMillan, 2005; Putnam & Borko, 1997).

Teachers found a level of meaning and usefulness from the professional

experiences that they mentioned, and because of that, they felt empowered to turn and use

the newly acquired knowledge in their classroom (Darling-Hammond, 2008). All six of

the educators to one degree or another referenced that their professional decisions were

most influenced by experiences where they had opportunities to: 1) participate within the

core discipline and reflecting upon that work; 2) engage in inquiry with other teachers

that focuses on improved classroom practice; and 3) engage in professional discourse

around theory-based research (Darling-Hammond, 2008; Marlink & Wahleithner, 2011;

Stokes, 2010).

Implications for Education

The findings from this study shed light on the Texas Writing Pilot as well as

influences of teacher-decision making and professional learning when scoring student

writing. The insights from this study provide shed light on possible implications from

teachers, administrators, and policy makers seeking to utilize authentic assessment for the

purposes of high-stakes assessment. The following section explores possible implications

for teachers, administrators, and policy makers.

Implications for Teachers

As one who works with students directly every day, a teacher can have incredible

insight into the needs of her students as well as expertise in designing meaningful

instruction and assessment for her students. As an experienced expert, teachers can play

an important role as an expert from the field with influential voices advocating for the

needs of their students. The data from the document analysis revealed a clear disconnect between how educators teach writing and how the state assesses writing. These dueling forces are ultimately what lead to the disconnect within assessment, instruction, scoring, and calibration. As the state continues to consider assessment implications because of new state standards, as well as new and alternative ways to authentically assess what a student knows, teachers must take an active role in contributing their voices to the process.

By taking part in studies such as The Texas Writing Pilot, teachers were able to have their voices heard and specifically influenced the design of the assessment. Teachers can and should play a role of advocacy by finding opportunities to articulate the support structures needed from the state to better teach and support student mastery of the standards. Even at a district level, teachers can advocate for additional opportunities to collaborate and learn alongside one another to calibrate and more tightly align scoring even as a team.

Implications for Administrators

Administrators play an important role in the learning process for teachers. Through meaningfully designed professional learning programs, teachers are able to improve their craft and apply new strategies and techniques in the classroom. The participants from this study identified personal experience, professional learning, and mentorship as three influential components in their decision-making processes. For administrators, this sheds light on the fact that learning for teachers does not only occur

163

in formalized professional development training, rather teachers learn in a variety of settings. As administrators develop learning opportunities for teachers, consideration should be given to a range of learning settings. Regardless of where or when the learning takes place, these experiences should be designed as such to where teachers have opportunities to 1) participate within the core discipline and reflecting upon that work; 2) engage in inquiry with other teachers that focuses on improved classroom practice; and 3) engage in professional discourse around theory-based research (Darling-Hammond, 2008; Marlink & Wahleithner, 2011; Stokes, 2010).

Additionally, when teachers in this study reflected on learning that was most impactful in their decision-making process, they often pointed to on-going, sustained professional learning, such as a week-long Advanced Placement Summer Institute, a summer institute hosted by the National Writing Project rather than a one-off one-day professional development. In a school district, when budgets are tight, longer, more intensive trainings are often the first to get cut. In considering the potential direct impact professional learning could have on instruction, administrators should carefully consider and invest in those experiences educators found as most influential in their teaching.

Implications for Policymakers

Regardless of the inditement from the TEA that The Texas Writing Pilot would not be scalable state-wide for the purposes of high-stakes assessment, it is undeniable that teachers reported that their writing instruction was more intentional and focused because

164

of the pilot. Teachers also reported that students were writing more than ever before, and they could actually see the improvement in the students' writing.

Policymakers looking to make assessment decisions for our students cannot disregard the positive effects on instruction and students that The Texas Writing Pilot had. The findings from the study confirm that the process of scoring student writing is a complex interpretive process. The study revealed how teachers do not separate their role as a rater from their role as a teacher, and because of this, they are able to make meaningful connections between assessment and instruction. Because scoring writing is a complex process influenced by a number of factors, professional learning, training, and support for educators implementing alternative assessments should also be considered when crafting new assessment policies.

Implications for Action

A recurring theme of the reports released by the TEA revealed there was no significant interrater reliability between teachers and the ETS raters. Ultimately this is problematic if the state were ever to see the value in having teachers score their own students' essays for assessment purposes. For those of us who were involved in the pilot, we know though there was insufficient time and support for teachers to calibrate. From the document analysis, the lack of interrater reliability did not definitively demonstrate that teachers cannot rate well. However, the data did suggest that teachers do not have agreement or understanding about how the state expects the scoring rubric should be utilized. In each of the phases of data analysis, the findings confirm that the process of

165

scoring student writing is a complex process of interpreting evaluating, reviewing, and justifying one's score. It was through the analysis of the metacognitive strategies teachers use that the data revealed teachers have a connection to the writer through the writing, even when evaluating writing from students who are not their own students.

As previous research suggests, the data from this study also supports the idea that teachers do not apply a rubric to a piece of writing in a vacuum without further consideration of external factors such as the student's ability, or even other assessment designs (Jeong, 2015a; Lumley, 2002). Ultimately, what educators believe about teaching and learning can, in fact, directly influence their decisions (Gill & Hoffman, 2009; McMillan, 2005; Putnam & Borko, 1997). From the data in this study, I was able to conclude the factors that come into play when teachers are making scoring decisions most often draw from their own experiences that include 1) personal experience; 2) professional learning and inquiry; and 3) mentorship, or professional discourse.

For a state seeking to utilize teachers in the scoring of assessments for the purpose of high-stakes assessment consideration to the timing and design of training must also include the knowledge and skills an individual teacher brings to the scoring table. As previous research suggests, a mitigating factor for inter-rater reliability can be the training raters receive because it can drastically influence reliability on the part of the rater (Jeong, 2015a; Knoch et al., 2007). While the findings from the study conducted by the TEA indicate that teachers did not score reliably consistent enough for this type of assessment to scaled statewide (Texas Education Agency, 2017), the reports also indicate

166

that teachers, at most, were provided with a three-hour scoring calibration training (Texas Education Agency, 2017). In order for the state to continue their efforts of teachers scoring student writing, future scoring training should take into consideration conditions, teachers from the study identified as optimal conditions for professional development. Those conditions include 1) opportunities to study by participating in experiences within the core discipline and reflecting upon that work; 2) collective inquiry with other teachers that focuses on improved classroom practice; and 3) engagement in professional discourse around theory-based research (Darling-Hammond, 2008; Marlink & Wahleithner, 2011; Stokes, 2010).

Recommendations for Further Research

As presented in the literature review, research related to the influences of teachers scoring student writing is sparse when it comes to a view outside of inter-rater reliability. As the state continues to look to expand opportunities for alternative assessment types, such as The Texas Writing Pilot, more work is needed to be done in considering professional learning experiences that may influence or mitigate what an educator brings to the scoring decisions so that ultimately inter-rater reliability might be improved.

Additionally, the study provided insight as to how the expanded model of scorer cognition (Wolfe & Feltovich, 1994) may be updated to include other influential factors to a teacher's scoring decision-making process. Further study into updating the model would provide an updated framework reflective of additional influences of a teacher's decision-making processes when scoring. This additional research could then potentially

167

lay the groundwork for additional research to consider how this model of scorer cognition could be used to influence professional learning or even pre-service teacher training and development.

Concluding Remarks

If states such as Texas wish to continue to find ways to strike a balance between assessment for accountability and assessment for learning, the role the teacher plays in the assessment process must be further considered and supported. In order for the states to view assessments such as The Texas Writing Pilot as a viable alternative assessment to the current assessment and accountability systems, writing teachers must be able to score student writing using a standardized rubric. As the role of teacher experience and background play in making scoring decisions (Jeong, 2015a) is further considered, it only stands to reason that new methods for training and supporting teacher education will be discovered and supported. The identification of the various factors that come into play when teachers are making scoring decisions provided insight into considerations for the next steps in the further development of alternative state assessments. Personal experiences, professional learning and inquiry, and, mentorship, or professional discourse must be the foundation of learning systems that support teacher's effective evaluation of student writing. The findings of this study should help inform state leaders and educators about the necessary considerations to support teachers further as they actively participate in the assessment process by scoring student writing for high-stakes assessment.

REFERENCES

Abbott, A. L. (2016). Locally developed performance assessments: One state's decision

to supplant standardized tests with alternative measures. *Journal of Organizational*

& Educational Leadership, *2*(1), 1–33.

Ackermann, E. K. (2004). Constructing knowledge and transforming the world. In

Tokoro M. and Steels L. (Ed.), *A learning zone of one's own: Sharing*

representations and flow in collaborative learning environments (pp. 15–37).

Washington, DC: IOS Press.

Act, N. C. L. B. (2002). No child left behind act of 2001. *Publ. L*, 107-110.

Barohny, E., & Heining-Boynton, A. L. (2007). Impact of an English-as-a-second-

language professional development program. *The Journal of Educational Research*,

101(1), 36–48. https://doi.org/10.3200/JOER.101.1.36-49

Behizadeh, N., & Engelhard, G. (2011). Historical view of the influences of measurement

and writing theories on the practice of writing assessment in the united states.

Assessing Writing, *16*(3), 189–211. https://doi.org/10.1016/j.asw.2011.03.001

Birgin, O., & Baki, A. (2007). The use of portfolio to assess student's performance.

Journal of Turkish Science Education, (4), 75-90.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in*

Education, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain.

Educational Researcher, *33*(8), 3–15. https://doi.org/10.3102/0013189x033008003

Borko, H., Shavelson, R. J., & Stern, P. (1981). Teachers' decisions in the planning of

    reading instruction. *Reading Research Quarterly*, *16*(3), 449–466.

    https://doi.org/10.2307/747411

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative*

    Research Journal, *9*(2), 27.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative*

    Research in Psychology, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bruner, J. S. (1990). *Acts of meaning*. London: Harvard University Press.

Causey-Bush, T. (2005). Keep your eye on Texas and California: A look at testing,

    school reform, no child left behind, and implications for students of color. *The*

    Journal of Negro Education, *74*(4), 332–343.

Cimbricz, S. (2003). State-mandated testing and teachers' beliefs and practice.

    Educational Policy Analysis Archives, *10*(2), 121.

    https://doi.org/epaa.asu.edu/epaa/v10n2.html

Clark, C. M., & Peterson, P. L. (1984). Teachers' thought processes (Occasional Paper

    No. 72). East Lansing, MI: Institute for Research on Teaching, Michigan State

    University.

Coffman, W. E. (1971). On the reliability of ratings of essay examinations in English.

    Research in the Teaching of English, *5*(1), 24–36.

Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper & L. Odell (Eds.),

Evaluating writing: Describing, measuring, judging (pp. 3–31). Urbana, Illinois: National

Council of Teachers of English.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five*

approaches. Thousand Oaks, CA: Sage.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the*

research process. Thousand Oaks, CA: Sage.

Cunningham, D., & Duffy, T. (1996). Constructivism: implications for the design and

delivery of instruction (pp. 131). In D. H. Jonassen (Ed.), *Handbook of Research for*

Educational Communications and Technology. New York, NY: Routledge.

Darling-Hammond, L. (2008). Teacher learning that supports student learning. In B. Z.

Presseisen (Ed.), *Teaching for intelligence* (pp. 91–100). Thousand Oaks, CA: Sage.

Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of*

performance assessment in achieving 21st century standards of learning. Stanford,

CA: Stanford Center

Dempsey, M. S., PytlikZillig, L. M., & Bruning, R. H. (2009). Helping preservice

teachers learn to assess writing: practice and feedback in a web-based environment.

Assessing Writing, *14*(1), 38–61. https://doi.org/10.1016/j.asw.2008.12.003

Diederich, P. B., French, J. W., & Carlton, S. T. (1961). Factors in judgments of writing

ability. *ETS Research Bulletin Series*, *1961*(2), i–93. https://doi.org/10.1002/j.2333-

8504.1961.tb00286.x

Diederich, P. D. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Division of Research and Analysis Office of Academics. (2018). *Enrollment in Texas public schools 2017-18*. Texas Education Agency.

DuFour, R., & DuFour, R. (2013). *Learning by doing: A handbook for professional learning communities at work*. Bloomington, IN: Solution Tree Press.

Duran, D., & Tipping, K. (2017). *Learning by teaching: Evidence-based strategies to enhance learning in the classroom*. New York, NY: Routledge.

Dutro, E., Selland, M. K., & Bien, A. C. (2013). Revealing writing, concealing writers: High-stakes assessment in an urban elementary classroom. *Journal of Literacy Research: JLR*, *45*(2), 99–141. https://doi.org/10.1177/1086296X13475621

Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. New York, NY: Peter Lang Inc.

Ellison, S. (2012). Intelligent accountability: Re-thinking the concept of "accountability" in the popular discourse of education policy. *Journal of Thought*, *47*(2), 19–41. https://doi.org/10.2307/jthought.47.2.19

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing naturalistic inquiry: a guide to methods*. London: Sage.

Flick, U. (2009). *An introduction to qualitative research*. London: Sage.

Gibson, W., & Brown, A. (2009). *Working with qualitative data*. London: Sage.

Gill, M. G., & Hoffman, B. (2009). Shared planning time: A novel context for studying teachers' discourse and beliefs about learning and instruction. *Teachers College Record*, *111*(5), 1242–1273.

Gomez, E. L. (1999). *Assessment portfolios and English language learners: frequently asked questions and a case study of Brooklyn International High School*. Northeast and Islands Regional Educational Laboratory, Brown University.

Griffith, R., Massey, D., & Atkinson, T. S. (2013). Examing the forces that guide teaching decisions. *Reading Horizons*, *52*(4), 2.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In *Handbook of qualitative research* (vol. 2, pp. 105–117). Thousand Oaks, CA: Sage.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development. *Phi Delta Kappan*, *90*(7), 495–500. https://doi.org/10.1177/003172170909000709

H.B. 1164, 84th Leg. R.S. (Tex. 2015).

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing*, *8*(1), 5–16. https://doi.org/10.1016/S1075-2935(02)00029-6

Harel, I., & Papert, S. (1991). *Situating constructionism*. Norwood, NJ: Ablex Publishing Corporation.

Hubbard, R. S., & Power, B. M. (1993). *The art of classroom inquiry*. Portsmouth, NH: Heinemann.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and

    prevailing trends. *Review of Educational Research*, *60*(2), 237–263.

Huot, B. (2002). *(Re)articulating writing assessment for teaching and learning*. Logan,

    UT: Utah State University Press.

Huot, B., O'Neill, P., & & Moore, C. (2010). A usable past for writing assessment.

    College English, *72*(5), 495–517.

Jeong, H. (2015a). Rubrics in the classroom: Do teachers really follow them? *Language

    Testing in Asia*, *5*(1), 6. https://doi.org/10.1186/s40468-015-0013-5

Jeong, H. (2015b). What is your teacher rubric? extracting teachers' assessment

    constructs. *Practical Assessment, Research & Evaluation*, *20*(6), 1–13.

John-Steiner, V. (1985). *Notebooks of the mind*. Albuquerque, NM: University of New

    Mexico Press.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and

    educational consequences. *Educational Research Review*, *2*(2), 130–144.

    https://doi.org/10.1016/j.edurev.2007.05.002

Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales.

    Language Testing, *26*(2), 275–304. https://doi.org/10.1177/0265532208101008

Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How

    does it compare with face-to-face training? *Assessing Writing*, *12*(1), 26–43.

    https://doi.org/10.1016/j.asw.2007.04.001

Knowles, M. S. (1970). *The modern practice of adult education*. New York, NY: Association Press.

Knowles, M. S. (1996). Andragogy: an emerging technology for adult learning. In *education for adults: Volume I: Adult learning and education*. London: Routledge.

Koretz, D., Stecher, B., & Deibert, E. (1992). *The Vermont portfolio assessment program: Interim report on implementation and impact, 1991-92 school year CSE technical report 350*. Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (2005). The Vemont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, *13*(3), 5–16. https://doi.org/10.1111/j.1745-3992.1994.tb00443.x

Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus and Giroux.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newburry Park, CA: Sage.

Lloyd-Jones, R. (1977). Primary trait scoring. In C. R. Cooper & L. Odell (Eds.), *Evaluating Writing: Describing, Measuring, Judging* (pp. 33–68). Urbana, IL: National Council of Teachers of English.

Loeb, S., & Hough, H. (2016). *Making use of waivers under ESSA*. Evidence Speaks Reports - The Brookings Institution.

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL. University: of Chicago Press.

Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, *19*(3), 246–276.

Marlink, J., & Wahleithner, J. (2011). *Improving students' academic writing: Building a bridge to success*. California Writing Project. Retrieved from https://lead.nwp.org/knowledgebase/improving-students-academic-writing-building-a-bridge-to-success/

Marshall, C., & Rossman, G. (2014). *Designing qualitative research*. Thousand Oaks, CA: Sage.

McMillan, J. H. (2005). Understanding and improving teachers' classroom assessment decision-making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34–43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x

Mc Millan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading practices decision-making*. Metropolitan Educational Research Consortium. Retrieved from https://eric.ed.gov/?id=ED447195

Merriam, S. (2009). *Qualitative research: a guide to design and research*. *San Francisco, CA*: John Wiley & Sons.

Merriam, S. B. (1998). *Qualitative research and case study applications in education. Revised and expanded from "Case study research in education."* San Francisco, CA: Jossey-Bass.

Merriam, S. B. (2001). Andragogy and self-directed learning: Pillars of adult learning theory. *New Directions for Adult and Continuing Education*, *2001*(89), 3–14. https://doi.org/10.1002/ace.3

Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M. (2012). *Learning in adulthood: A comprehensive guide*. San Francisco, CA: Jossey-Bass.

Montgomery, D. (2013, March 26). Texas house passes bill to reduce end-of-course exams in public schools | Fort Worth Star-Telegram. Retrieved August 11, 2018, from https://www.star-telegram.com/news/local/education/article230177534.html

Moore, C., O'Neill, P., & Huot, B. (2009). Creating a culture of assessment in writing programs and beyond. *College Composition and Communication; Urbana*, *61*(1), W107–W132.

Moss, P. A. (1994). Validity in high stakes writing assessment: Problems and possibilities. *Assessing Writing*, *1*(1), 109–128 https://doi.org/10.1016/1075-2935(94)90007-8

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, *46*(2), 532–566. https://doi.org/10.3102/0002831208328088

No child left behind act of 2001: Qualifications for teachers and professionals. (2008). (No. 20 U.S.C. § 6319) (Vol. 20).

Obama, B. (2009, March). *A complete and competitive American education*. Presented at

the Remarks of the President to the United States Hispanic Chamber of Commerce,

Washington Marriott Metro Center, Washington D.C. Retrieved from

https://obamawhitehouse.archives.gov/the-press-office/remarks-president-united-

states-hispanic-chamber-commerce

O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language

learners: Practical approaches for teachers*. Boston, MA: Addison-Wesley

Publishing Company.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership:

Journal of the Department of Supervision and Curriculum Development, N.E.A*,

49(8), 14–19.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York, NY:

Basic Books, Inc.

Papert, S. (2011, October 23). Whose fingers on the button? Retrieved from

https://www.independent.co.uk/news/education/education-news/whose-fingers-on-

the-button-1158688.html

Papert, S. (2006). Minding change. *Human Development*, *49*(4), 239–247.

https://doi.org/10.1159/000094373

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921–958. https://doi.org/10.3102/0002831207308221

Piaget, J. (2014). *Language and thought of the child: Selected works vol 5*. New York, NY: Routledge.

Picard, R. W., Papert, S., Bender, W., & Blumberg, B. (2004). Affective learning -- a manifesto. *BT Technology Journal*, *22*(4), 253.

Putnam, R. T., & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *International Handbook of Teachers and Teaching* (pp. 1223–1296). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-4942-6_30

Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture, and Activity*, *1*(4), 209–229. https://doi.org/10.1080/10749039409524673

Ross, J., & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of randomized field trial. *The Journal of Educational Research*, *101*(1), 50–60. https://doi.org/10.3200/JOER.101.1.50-60

Sacks, P. (2000). Standardized minds: The high price of America's testing culture and what we can do to change it. *National Association of Secondary School Principals. NASSP Bulletin*, *84*(616), 118.

Saddler, B., & Andrade, H. (2014). The writing rubric. *Educational Leadership*, *62*(2), 48–52.

Shavelson, R. J. (1973). What is the basic teaching skill? *Journal of Teacher Education*, *24*(2), 144–151. https://doi.org/10.1177/002248717302400213

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Smagorinsky, P. (1994). *Speaking about writing: Reflections on research methodology*. Thousand Oaks, CA: Sage.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessment. *Assessment in Education: Principles, Policy & Practice*, *5*(3), 335–351. https://doi.org/10.1080/0969595980050303

Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Stiggins, R. J. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practice*, *6*(3), 33–42. https://doi.org/10.1111/j.1745-3992.1987.tb00507.x

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758. https://doi.org/10.1177/003172170208301010

Stokes, L. (2010). The national writing project: Anatomy of an improvement

infrastructure. In Coburn C & Stein (Ed.), *Research and practice in education:*

Building alliances, bridging the divide (pp. 147–163). Lanham, MD: Rowman &

Littlefield Publishers.

Texas Education Agency. (2016a). Request for applications: The 2016-2018 writing pilot

program. Retrieved from https://tea.texas.gov/student-assessment/testing/staar/the-

texas-writing-pilot-program

Texas Education Agency. (2016b). SEC 39.02301 Writing assessment study; pilot

program. Retrieved from https://tea.texas.gov/student-assessment/testing/staar/the-

texas-writing-pilot-program

Texas Education Agency. (2016c). House bill 1164 writing pilot program: Report to the

governor and the Texas legislature. Retrieved from https://tea.texas.gov/student-

assessment/testing/staar/the-texas-writing-pilot-program

Texas Education Agency. (2017). House bill 1164 writing pilot program: report to the

governor and the Texas legislature. Retrieved from https://tea.texas.gov/student-

assessment/testing/staar/the-texas-writing-pilot-program

Texas Education Agency, (2018). House bill 1164 writing pilot program: Report to the

governor and the Texas legislature. Retrieved from https://tea.texas.gov/student-

assessment/testing/staar/the-texas-writing-pilot-program

Texas Education Agency. (n.d.). The Texas Writing Pilot program. Retrieved from

   https://tea.texas.gov/student-assessment/testing/staar/the-texas-writing-pilot-

   program

Tierney, R., Carter, M., & Desai, (1991). *Portfolio assessment in the reading-writing*

   classroom. Norwood, MA: Christopher-Gordon Publishers, Inc.

Tung, R. (2010). Including performance assessments in accountability systems: A review

   of scale-up efforts. Retrieved from Nellie Mae Education Foundation

   https://www.nmefoundation.org/getmedia/b9d7909f-b76d-4a4c-94d5-

   6079b8d110f9/cce-exec-summary-final?ext=.pdf

Turkle, S., & Papert, S. (1990). Epistemological pluralism: Styles and voices within the

   computer culture. *Signs: Journal of Women in Culture and Society*, *16*(1), 128–157.

   https://doi.org/10.1086/494648

U.S. Department of Education, Office of Elementary and Secondary Education. (2015).

   Fundamental change: Innovation in America's schools under race to the top.

   Retrieved from https://www2.ed.gov/programs/racetothetop/rttfinalrpt1115.pdf

Valdéz Pierce, L., & O'Malley, J. M. (1992). Performance and portfolio assessment for

   language minority students. *NCBE Program Information Guide Series.*

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological*

   processes. Cambridge, MA: Harvard University Press.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. New York, NY: Cambridge University Press.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, *70*(9), 703–713.

Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan*, *75*(3), 200–208.

Wiggins, G. (1996). Practicing what we preach in designing authentic assessments. *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A*, *54*(4), 18.

Williamson, M. (1994). The worship of efficiency: Untangling theoretical and practical considerations in writing assessment. *Assessing Writing*, *1*(2), 147-73. https://doi.org/10.1016/1075-2935(95)90021-7

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, *24*, 173–209. https://doi.org/10.2307/1167270

Wolfe, E. W., & Feltovich, B. (1994). Learning to rate essays: A study of scorer cognition. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, *50*(3), 483–503. https://doi.org/10.2307/358862

Zhang, J. (2016). Same text different processing? Exploring how raters' cognitive and

meta-cognitive strategies influence rating accuracy in essay scoring. *Assessing*

Writing, *27*, 37–53. https://doi.org/10.1016/j.asw.2015.11.001

Zmeyov, S. I. (1998). Andragogy: Origins, developments and trends. *International*

Review of Education, *44*(1), 103–108. https://doi.org/10.1023/A:1003107931006

APPENDIX A

Initial Recruitment Email

Initial Recruitment Email

Good Morning,

My name is Heather Cato. I am a graduate student in the Department of Reading at Texas Woman's University. I am conducting research to explore the process teachers use in decision-making when scoring student work samples and to consider the teacher knowledge and professional learning that contribute to the different scoring approaches teachers use. My goal is to interview and observe six educators in the North Texas area near or around Dallas-Fort Worth with experience teacher writing at the fourth grade, seventh grade, or ninth grade level. All Texas Woman's University expectations for safe research will be followed. I hope that you will allow me this wonderful opportunity to interview you and benefit from your experiences. If you are willing to participate, please respond to this email by providing me with your email and phone number. While there is a risk of loss of confidentiality in all email, downloading, and Internet transactions, *Confidentiality will be protected to the extent that is allowed by law.* Because participation is voluntary, you may withdraw at any time.

Should you need to contact me, I can be reached at hcato@twu.edu or 817-706-9955.

APPENDIX B

Online Participant Questionnaire

Online Participant Questionnaire

The purpose of this questionnaire is for gathering background information regarding your experiences and specialized training as it relates to writing scoring and instruction. The questions were designed to be open ended so you can respond as you see fits. If you have any questions, please do not hesitate to ask.

1. How would you describe yourself as a writing teacher?

2. How would you describe yourself as a writer?

3. As it relates to writing instruction, what experiences or specialized training have you had?

4. How many years have you taught?

5. What grade levels have you taught?

6. Describe any experience you have had with teaching or preparing students for a state writing assessment?

7. What teaching certifications do you have?

APPENDIX C

Pre-Interview Protocol

Pre-Interview Protocol

Thank you for taking the time to meet with me. The purpose of this interview is for gathering a bit more information regarding your experiences and specialized training as it relates to writing scoring and instruction. The questions were designed to follow up on some of your responses in the questionnaire as well as go deeper into your beliefs and practices regarding writing instruction and assessment. As you answer each question, I may have a follow up question or two in order to gain a more complete picture of your initial answer.

- As it relates to writing instruction, what experiences or specialized training have you had?

- Have you ever taught or prepared students for a state writing assessment?   If yes, what training or experiences have you had that specifically have prepared you to prepare students for state writing assessments?

- In what ways do you think your beliefs about writing influence your instruction?

- Please describe how writing instruction typically looks in your classroom

- You mentioned _____ training, in question 1, please describe any take aways you had from that training and how you used that information in your classroom.

APPENDIX D

First Think Aloud Protocol

First Think Aloud Protocol

I am now going to ask you to rate a set of student writing essays. Remember, I am conducting a study to explore teacher decision-making when scoring student work samples and to consider the teacher knowledge and professional learning that contribute to the different scoring approaches teachers use. For these essays, I would now like you to talk and think aloud as you review these papers. First, you should identify each essay by the ID number at the top of the page. Then, as you review each essay, you should vocalise your thoughts, and explain what you notice about the writing and how you might score this essay if you were to give it a grade. For this exercise, I will not be providing you with a specific rubric, so you will need to draw upon your knowledge and prior experiences in order to inform your decision about how you would score each paper. As you are reviewing and evaluating each paper, it is important that you keep talking the entire time and vocalizing your thoughts all the time. In order to make sure there are no lengthy silent pauses in your rating, I propose to sit here, and prompt you to keep talking if necessary. I will sit here while you rate and talk. I will say nothing more than give you periodic feedback such as `mhm', although I will prompt you to keep talking if you fall silent for more than 10 seconds.

APPENDIX E

Second Think Aloud Protocol

Second Think Aloud Protocol

I am now going to ask you to rate a set of student writing essays. Remember, I am conducting a study to explore teacher decision-making when scoring student work samples and to consider the teacher knowledge and professional learning that contribute to the different scoring approaches teachers use. For these essays, I would now like you to talk and think aloud as you review these papers. This time I have provided you with a writing rubric I would like for you to use in order to score each essay.

As before, I would like you to talk and think aloud as you rate each essay, while this tape recorder records what you say. First, you should identify each essay by the ID number at the top of the page. Then, as you review each essay, you should vocalize your thoughts, and explain what you notice about the writing what you are considering as you are making a decision about what score to give the paper. You will give each paper a score for each of the domains of the rubric. Remember, as you rate each paper, you should vocalize your thoughts, and explain why you give the scores you give. As you are reviewing and evaluating each paper, it is important that you keep talking the entire time and vocalizing your thoughts all the time. In order to make sure there are no lengthy silent pauses in your rating, I propose to sit here, and prompt you to keep talking if necessary. I will sit here while you rate and talk. I will say nothing more than give you periodic feedback such as `mhm', although I will prompt you to keep talking if you fall silent for more than 10 seconds.

APPENDIX F

Texas Writing Pilot Rubric
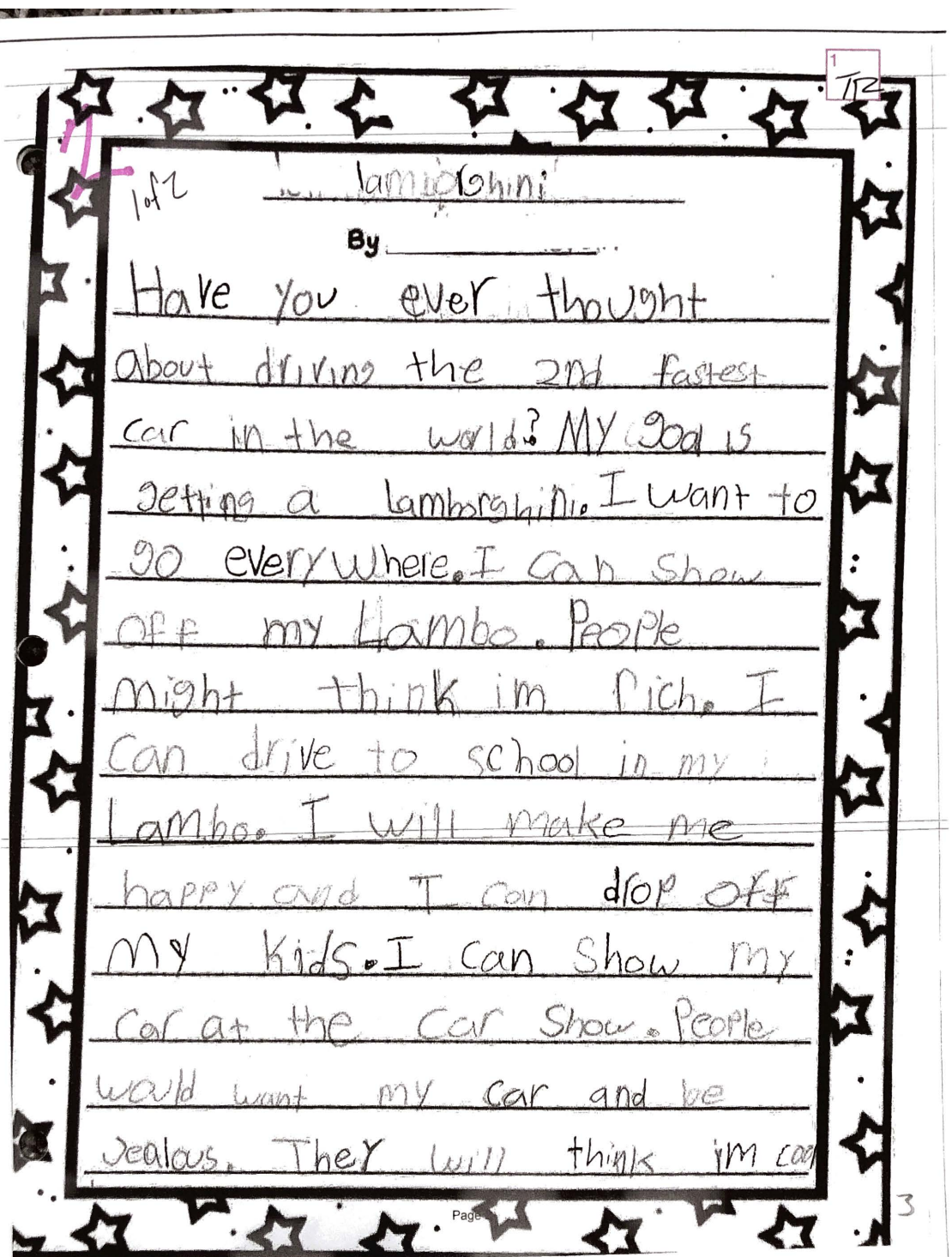
Texas Writing Pilot Rubric

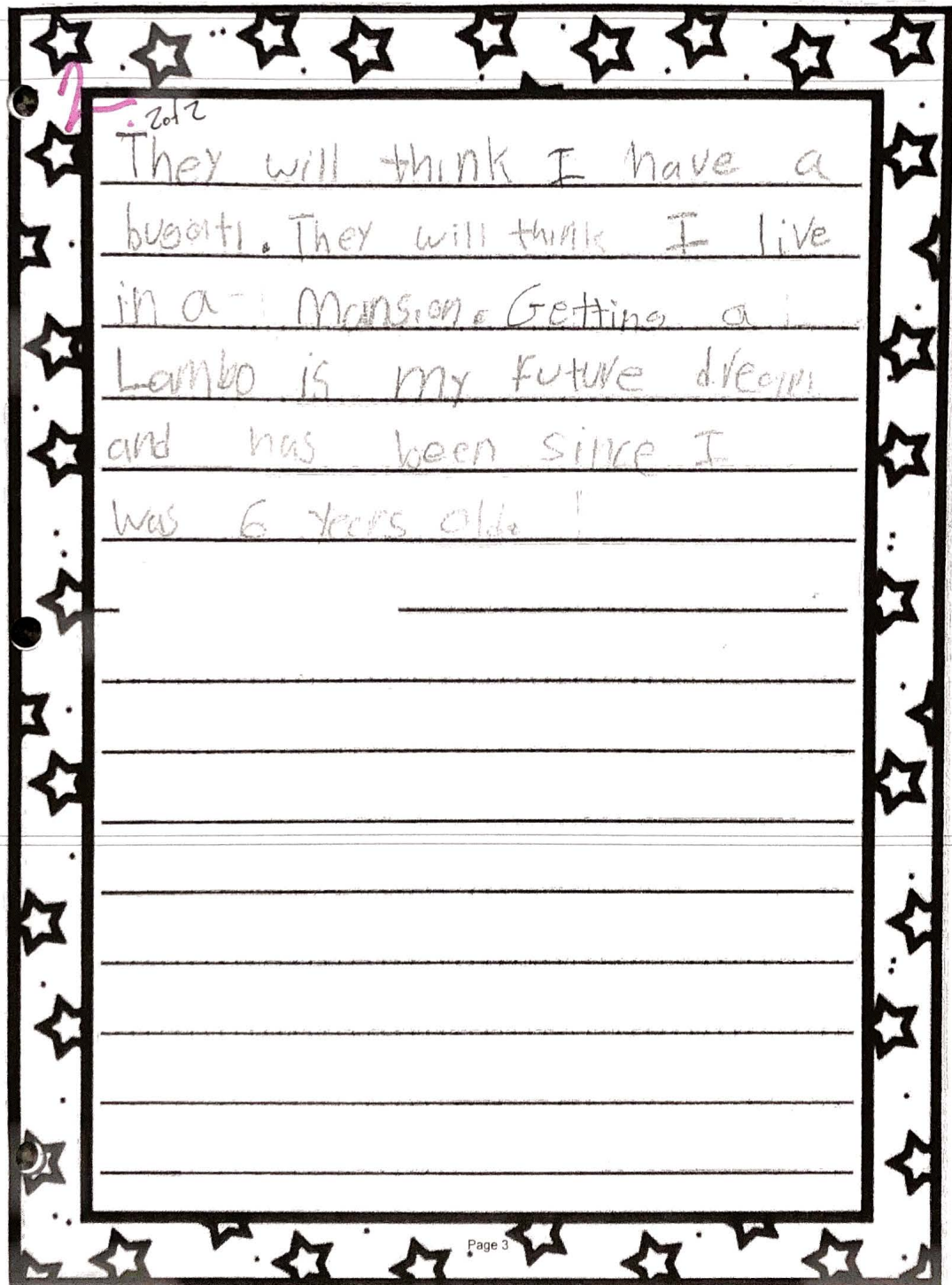| Very Limited	Limited	Basic	Satisfactory	Accomplished	Exceptional
1	2	3	4	5	6
ORGANIZATION: STRUCTURE, FOCUS, AND PROGRESSION					
The composition does not include a central idea, thesis, or theme. The composition lacks an organizational structure. The composition lacks a central focus and is therefore incoherent and not unified. The composition includes no evidence of connections between ideas.	The composition includes a central idea, thesis, or theme that is mostly unclear. An organizational structure may be evident, but it does not support the development of the central idea, thesis, or theme. The focus is inconsistent, causing the composition to be mostly incoherent and not unified. The sentences, paragraphs, and/or ideas are not clearly connected. Repetition of ideas causes serious disruptions in the flow of the essay.	The central idea, thesis, or theme is somewhat clear. The organizational structure only minimally supports the development of the central idea, thesis, or theme. The focus is at times inconsistent, causing lapses in the composition's coherence and unity. The sentences, paragraphs, and/or ideas are connected by mechanical, formulaic transitions. Some repetition of ideas causes minor disruptions in the flow of the essay.	The central idea, thesis, or theme is clear. The organizational structure is appropriate and adequately supports the development of the central idea, thesis, or theme. The focus is generally consistent and clear, helping the composition remain mostly coherent and unified. The sentences, paragraphs, and/or ideas are connected by logical and mostly effective transitions.	The central idea, thesis, or theme is clear and skillfully presented. The organizational structure is appropriate and effectively supports the development of the central idea, thesis, or theme. The focus is consistent and clear throughout, contributing to the composition's sustained coherence and unity. The sentences, paragraphs, and/or ideas are connected by logical, effective transitions.	The central idea, thesis, or theme is clear and thoughtful. The organizational structure enhances the development of the central idea, thesis, or theme. The focus is consistent and clear throughout, contributing to the composition's sustained coherence and unity. The sentences, paragraphs, and/or ideas are connected by purposeful, logical, and highly effective transitions.
1	2	3	4	5	6
CONTENT: SUPPORT AND ELABORATION					
The composition includes few, if any, details and/or examples related to the topic or theme. The composition may be too brief to reflect an understanding of the writing purpose and/or communicate the writer's intent.	The composition includes details and examples that are list-like and/or too vague to adequately develop the topic or theme. The composition reflects an inadequate understanding of the writing purpose and/or is unable to communicate the writer's intent.	The composition includes mostly relevant details and examples, but they are too general or partially presented to adequately develop the topic or theme. The composition reflects some understanding of the writing purpose and/or only somewhat communicates the writer's intent.	The composition includes relevant details and examples that adequately develop the topic or theme. The composition reflects an adequate understanding of the writing purpose and/or adequately communicates the writer's intent.	The composition includes relevant, specific details and examples that clearly develop the topic or theme. The composition reflects a thorough understanding of the writing purpose and/or strongly communicates the writer's intent.	The composition includes details and examples that are specific, well chosen, relevant, and enhance the development of the topic or theme. The composition reflects a thorough and insightful understanding of the writing purpose and/or clearly communicates the writer's intent in ways that are original and thoughtful.

Texas Writing Pilot Rubric

1	2	3	4	5	6
LANGUAGE					
The composition includes limited diction that is frequently used incorrectly and does not contribute to creating an appropriate/effective tone and style. Literary and/or rhetorical devices are typically missing. The composition includes sentences that are mostly unclear and illogical. Sentences are choppy, irregular, awkward, or incomplete and do not establish the relationships among ideas.	The composition includes simplistic diction that only minimally contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, do not contribute to the quality or effectiveness of the composition. The composition includes sentences that are at times unclear and illogical. Sentences are mostly simple, may include inappropriate fragments, and may not establish the relationships among ideas.	The composition includes sometimes vague or general diction that inconsistently contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are somewhat effective in contributing to the quality or effectiveness of the composition. The composition includes sentences that are mostly clear and logical. Sentences and phrases may at times be awkward or only somewhat controlled, occasionally weakening the relationships among ideas.	The composition includes mostly appropriate diction that satisfactorily contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are effective and contribute to the quality or effectiveness of the composition. The composition includes sentences that are consistently clear and logical. Sentences and phrases are adequately controlled and usually establish the relationships among ideas.	The composition includes specific diction that consistently contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are engaging, and contribute to the quality or effectiveness of the composition. The composition includes sentences that are consistently clear, logical, and varied in structure. Sentences and phrases are skillfully controlled and effectively establish the relationships among ideas.	The composition includes purposeful and precise diction that strongly contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are effective, engaging, original, and enhance the quality or effectiveness of the composition. The composition includes sentences that are consistently clear, logical, and varied in structure. Sentences and phrases are sophisticated in construction and strongly establish the relationships among ideas.

2	4	6
CONVENTIONS		
The composition includes a variety of errors reflecting limited or no control of basic writing conventions (spelling, capitalization, punctuation, grammar, and usage). The composition may require extensive editing for conventions errors or may be too brief to evaluate for control of conventions. The composition demonstrates limited or no control of sentence boundaries. If included, paragraph breaks interfere with meaning or demonstrate only a basic understanding of their use.	The composition demonstrates sufficient control of standard writing conventions (spelling, capitalization, punctuation, grammar, and usage). The composition may require minor to moderate editing for conventions errors. The composition demonstrates reasonable control of sentence boundaries. If included, paragraph breaks demonstrate adequate understanding of their use.	The composition demonstrates consistent command of standard writing conventions (spelling, capitalization, punctuation, grammar, and usage). The composition requires minor, if any, editing for conventions errors. The composition may contain purposeful manipulation of conventions for effect. The composition demonstrates consistent control of sentence boundaries, enhancing the composition. If included, paragraph breaks are well controlled and purposeful.

196

APPENDIX G

Grade 4 Student Writing Samples

lamborghini

By _____

Have you ever thought about driving the 2nd fastest car in the world? My goal is getting a lamborghini. I want to go everywhere. I can show off my Lambo. People might think im rich. I can drive to school in my Lambo. I will make me happy and I can drop off my kids. I can show my car at the car show. People would want my car and be jealous. They will think im cool

2012

They will think I have a
bugaiti. They will think I live
in a Mansion. Getting a
Lambo is my future dream
and has been since I
was 6 years old!

199

November, 21 2016
final Rhagh Draft

My Life in Summer

To know my opinion, is   introduction
when I Be a Pepper in
may life and having fun.

fistly, summer is so cool   P1
and you can take out all   2
the coldness. When summer
comes it will refesh everyBody.
You can say that, summer can
Be spicy hot But it's more cool
and you can go swimming.
Summer will Be the Best one to
go have some fun.

   Lastly, summer can Be hot P2 3
But try doing D.i.y. When summer
it's the Best Season to hangout
and Be a hot spicy pepper.
Summer was a hot season But
try to get over with it, summer
is so hot But when you see something   64
try to make some cool stuff.

3

conclusion

① Overall, summer will always 4 Be in my Life and I can go have an a anvanture with my firends or family together.

65

201

roller coaster oof tear!!!

I was at six Flage's it was pich black. I was about to ride shock wave the scaryest roller coaster in six Flag's well at least to me.

I was tired of wating in the long line well I geuss it is good because I was scared, nerves and exited all at the same time.

It was finally my turn to ride shock wave me and my dad got in the frist lether seat.

Then I said to my day "I don't want to ride shock wave any more" then he said "you will be fine" then Ite went in loop's and loop's (AhAh) finally the roller coaster was over (yay!) then my dad said "I told you that you will be fine.

But I think my face is green

4

because I felt like I
will throw up.
    Then I thought to my
self now I know why shock the
wave is the scargest rollercoaster
I won't be riding shock wave
any time soon!!!

4

Writing Pilot Composition Page

"Gulp" I got on to the stage, I glanced at all the people in front of me cheering, smiling, and shouting. I said to myself, "I can do this." I walked a litte closer to the front of the stage and said, "It's time." I could picture myself on the red carpet walking and posing, and I thought to myself I can do this. I held the microphone close to my cherry lips, it was dead silence. I started singing, when I started singing a smile grew bigger and bigger on my face, my soul filled with joy and happiness. My voice fell. I was completley done with my song. I've never felt that much joy in my heart till now. After all the contestants were done, they were going to announce the winner. My heart started pumping rapidly, "Thump, Thump, Thump, Thump, Thump, Thump," the anounncer anounced the winner, "The winner is _____." I broke into tears of joy. Till then I was never shy or nervous to sing in front of people. I infer to people who are shy or nervous get out there and sing your heart out, but I also question myself why was I ever afraid to sing in front of people? Maybe my heart will tell me.

Rollercoaster Of a Lifetime

On a hot, sunny day, me and my family went to Sixflags and we were going up the stairs to the green rollercoaster called the Shockwave. It had two scary looking loops. It was my first time to ever go on the ride and I was terrifide.

We were at the front of the line waiting nervesly at the gait. The gate was metal and long, and soon it would open and I would have to get on.

I got onto the green cart and it started to move "Ahhhhhh!" The carts left and we went down the fast slope.

We went around the terrifing loops and up and down. Then we went over the cloudy blue lake. Finaly we were jerked back and, WE STOPED!

I thought it would never end, but I guess, I still had fun. "Again, Again, Again" I yelled.

Writing Pile

Dear Mr. ___ May we bring our bikes to
school. It is helty for you and you will
use muscles and you will have less
energy.

You can work hader becese
you are tiered and you can focas
better and get better grades.

You can execise better and
you can be streng and helty.

206

APPENDIX H

Grade 7 Student Writing Samples

Phones Can Be Class Friendly!

Have you noticed that lots of schools in America don't allow phones in classrooms? Teachers may have some disputes about this subject, but they also need to recognize the advantages of how phones can be very useful in various classrooms. Students should be allowed to bring phones to class because they can use the phones to research various subjects, to sometimes use music to help concentrate, and even to contact their parents or guardians if needed.

There are lots of students who need to research academic information for a certain class. It's effortless to use your phone for these situations mainly because it's easy to access as well as looking up information better than a school provided computer will ever be. Researching on a school provided computer usually needs to have you logged in to the computer before accessing the websites you need to use. This is a problem itself since certain students may forget subjects such as usernames and passwords easily. Phones, however, can simply be accessed from the student's backpack as well as their pocket and you can get all the information you need with a push of a button.

Some students often have issues working with the noise and the conversations in the classroom environment and may find it easier to concentrate with music. Take under advisement that you can't modestly listen to any song you want. Lyrical songs are huge

distractions when working on classwork. If the student listens to songs like classical or soothing songs, they block the atmosphere's noises and conversations with a nice, calm, soothing tune while working on assignments in class.

There are also various students who need to contact their parents or guardians to ask for something they forgot or for an emergency. I one time needed to call my mother to ask who will pick me up from school that afternoon. Since I didn't have a phone at the moment, I had to walk to the office door, ring the buzzer, and wait for the office to unlock the door for me to access to the school provided phone. There usually is a small line and need to wait. If many students weren't allowed to use their devices in class, lots of students would need to interrupt their lesson would need to wait a huge line just to contact their parent or guardian. It would eventually lead to being a way for students to have an excuse to leave class. If students would have access to their devices in class, they could ask for permission to text whoever they need to contact at any time necessary.

Providing phones in the classrooms can be very useful in many different ways. To research, listen to music to help concentrate, and contact their parents or guardians when needed are three of many other positives that the phones can be as useful as an everyday school provided device. There may be some negatives, but there are also many positives of using these devices in classrooms that lots of school staff don't recognize. Phones can be helpful to the classroom environment and we should try to let the students use their phones in class to help make learning faster, easier, and a class friendly tool in various classrooms.

Student Name:_____

Prompt Number: ___3_____

Page __1__ of __1__

Writing Pilot Composition Page

The one Person that's important to me
    I have that special someone in my life that I can always count on. She could never be replaced for nothing in the world and that is my bestfriend Name ~~____~~

    My bestfriend and I have been through thick and thin. She is very important to me because she has never left my side for nothing in the world. We've always stuck together since fifth grade. We are like to peas in a pod. ____ is A person that will always have my back and that I can tell evrey thing too. I knew when I first met her that she was going to be the one that's going to be my side forever. We've only had one argreement and she still didnt leave my side. Sometime we get mad at each other but at the end of the day we are still bestfriends. She is really the only person that I can really talk to about certain things because I know that she can keep whatever I tell her. Every time she comes over or I go over house one of us has to tell something to one another.

    My bestfriend is the main one I trust. She could never be replace no matter what. I can always put my trust in her because I know that she will be the one to carry me through. I could never ask anyone else to be there for but her. She is like my bhad sister to me. She always said to me when I'm feeling down that "I love you more than anything I'll never let any one come and take your spot. And that the truth." Ilove my bestfriend She could never be traded for nothing that I really want. She will always be my bestfriend.

Texas Writing Pilot 2016-2017-Draft

Dump Those Phones

Being able to bring phones to class has always been a major issue throughout many schools. However, great these little machines may be, they are also the main cause of decreased focus and concentration in the classroom. Phones have proven many times to be a distraction to the student possessing the phone and those around them. Accordingly, phones should not be allowed in class because they draw the concentration of the class away from the discussion and can lead to decreased grades.

When students are given the privilege of having their phones in class, they eventually lose their concentration. Soon, they begin to pay more attention to the amount of time left in class then learning and comprehending the lesson. Although some may disagree and argue that teachers could take up the student's phones, the students still become anxious to play on their phones again, or text their family and friends. Also, phones in class have tended to change a student's priorities from studying and doing well in school to playing and texting on their phones. This has been proven by numerous researchers who report that 92% of students text with their phones during class. While these students are texting, they become distracted while they wait for a reply so they miss what the teacher is saying because they were focused on another subject.

In many schools, phones aren't allowed to be out during class. Due to this, many students attempt to sneak their phones, which is never an option or good thinking. While sneaking their phones, many people underestimate their teacher and eventually their phones are taken up. Once this happens, the students or their parents will have to pay in order to retrieve their phones. Of course some may say that the students could pay for themselves, but it is still the parents who provide them with money and it can be a

continuous burden for those who are tight on money. My friend   once ran into this conflict and became very overwhelmed. She had used her phone during class to text her parents, but she had forgotten to ask so her phone was taken up. Her family was in a very tight position so when the time came to pay the fine, her parents became very stressed and had to work extra. If phones weren't allowed in class, then this wouldn't have occurred in the first place.

As well as becoming a burden for most parents, cellphones in class can also become a serious distraction for the student who owns the phone and for the students around them. This distraction can become an encumbrance for others when the owner has to make an urgent phone call or text in the classroom. When a student has to text or call their parents during class, it becomes a major distraction and can lead to loss of concentration and focus throughout the class. Even if students were to step out of the classroom and make their phone call or text, it still becomes a distraction to those around when a person goes in and out of the room. Although there are a list of positives, the list of negatives definitely overrules. Most of these reasons can become life-changing to the student's future when they get to college, because these are the times that they need to concentrate the most. When phones are buzzing around the classroom, you will become unaware of what your professor is telling you. Therefore, you miss points on the test because you didn't have the correct notes. This occurred one day when my classmate sitting next to me got up to make a call. When she got up, I was distracted by the sudden movement and my focus for the teacher was neglected and instead was drawn to my classmate. At that very moment was when the teacher listed the most important facts on the review, and due to this, I missed those questions on the test.

Although phones are entertaining, they aren't always there for your benefit. They can be misleading and can lead you away from a successful future. So, why don't we dump those phones today?

3

● Name

Totem by Eamon Grennan

The poem Totem by Eamon Grennan was basically about a jack-o-lantern taking us through it's journey of going through death.

4
4
3
2

A pumpkin was put outside. Your probably asking,"Why was the pumpkin put out in the first place", and the answer is Halloween. In the beginning of the poem it said, "All Souls' over, the roast seeds eaten, I set on a backporch post our sculpted pumpkin under the weather, warm still for November." The line is just saying that "All Souls'" means November 2nd (Halloween), the point of view is from a pumpkin, the "backporch post" tells us the setting, "sculpted pumpkin" means that it's a carved pumpkin making it into a jack-o-lantern, "under the weather" is another way of saying feeling sick, and "warm still for November" means that as the sun beams down on it, you can conclude that the pumpkin is rotting faster by the minute. Another part of the poem is "it bows its boogeyman face of dread", the word 'boogeyman' is not a everyday poem word you would usually hear.

Overall the poem Totem was written to just tell us that eventually we will all reach the same fate, and we should just accept it and move on.

Thump! Thump! Thump! I heard footsteps dancing down the hall, creeping closer and closer like they were coming for me. Creeeek! My whole body started to shake. The door opened. "   is up next for the district speech competition," the frightful voiced announced. I gulped down my feelings of fear and cautiously stepped onto the stage to see the wide-eyed audience staring back at me, waiting for me to speak. As I walked to the podium my feet suddenly felt like they were loaded with weights. Although I was trembling, I managed to take a deep breath and begin my speech. Words felt like they were creeping out of my mouth so slowly at first, but then I took another breath and felt myself relax, helping me to deliver the rest of my speech more confidently. I took one last glance at the audience before I left the stage. Were they smiling at me?

I felt so relieved and happy with my performance, but those feelings did not last as I sat and listened to the speeches of the other competitors. My heart sank more and more as each person delivered awesome speech after awesome speech. The fear and doubt about my speaking ability returned. I began arguing with myself. "You did so badly," the voice in my head said. "Be quiet," I tried to argue back. To make things words, the judges informed us that the results from the speech competition wouldn't be revealed until later that evening. UGH! The anxiety of not knowing how I did dragged on me as I got into the car. "It's going to be okay," I heard my mom's comforting voice say as I slipped on my seat belt. "The important thing is that you tried," my dad added. I know my parents were trying to help, but I still felt the exact same way I did before I even stepped on the stage.

Once we got back home, I sat down at my desk, put my phone to the side, and tried to do my homework. I tried and tried to do my homework, but my mind kept wandering back to my speech. I replayed it over and over my head. "Did I speak clearly enough? Could everyone hear me? Did I speak with emotion and energy? Did I make good points?" I was driving myself nuts and was so thankful that mom interrupted my pity party and called me for dinner.

Just as we finished eating, the phone rang. My body went stiff, and I couldn't move a muscle…not even to reach for my phone. Dad reached over and picked up my phone. "Hello?" he answered. I heard nothing. Dad only nodded his head in response to the voice on the other end of the call. I held my breath, clenched my teeth, and closed my eyes. This one-minute call felt like pure torture. My dad finally hung up, stood face-to-face with me and took a slow, deep breath. "   I know you tried, so hard, but I have to let you know that YOU MADE IT TO THE REGIONALS!!!" At first my heart sank. Then, I realized what dad had just told me. I blinked my eyes, once, twice, and then screamed, "YAY!!" I almost burst into tears of happiness. Making it to regionals in speech was one of the most memorable events of my life and I will never forget it!

Student Name:

Prompt Number: 1

Writing Pilot Composition Page

4
4
4
2

When I grow up I want to be a
lawyer because lawyers make alot of
money and lawyers also help people.
Lawyers help people win cases and
that's what I want to do help people.
If I can't be a lawyer than I want
to be vetenerian because I can help cure
animals and I can see animals every
day all day. Vetenians and Lawyers are
totally different, I don't know what is
the amount of money the vetenerians
make, but It looks like a good or fun
job to work in, those two jobs are the
only jobs that I have interest in.
The only thing that I don't like about
being a lawyer is that It is a very
very long proccess which takes extremely
long to take care of. And one thing
that I don't like about being
a vetenerian is that I have to deal
with animals that die and that's really
sad, or sometimes I will have to
deal with animals that need shots
and some dogs can be really egressive when
it comes to shots.

APPENDIX I

Grade 9 Student Writing Sample

Phones Can Be Class Friendly!

Have you noticed that lots of schools in America don't allow phones in classrooms? Teachers may have some disputes about this subject, but they also need to recognize the advantages of how phones can be very useful in various classrooms. Students should be allowed to bring phones to class because they can use the phones to research various subjects, to sometimes use music to help concentrate, and even to contact their parents or guardians if needed.

There are lots of students who need to research academic information for a certain class. It's effortless to use your phone for these situations mainly because it's easy to access as well as looking up information better than a school provided computer will ever be. Researching on a school provided computer usually needs to have you logged in to the computer before accessing the websites you need to use. This is a problem itself since certain students may forget subjects such as usernames and passwords easily. Phones, however, can simply be accessed from the student's backpack as well as their pocket and you can get all the information you need with a push of a button.

Some students often have issues working with the noise and the conversations in the classroom environment and may find it easier to concentrate with music. Take under advisement that you can't modestly listen to any song you want. Lyrical songs are huge

distractions when working on classwork. If the student listens to songs like classical or soothing songs, they block the atmosphere's noises and conversations with a nice, calm, soothing tune while working on assignments in class.

There are also various students who need to contact their parents or guardians to ask for something they forgot or for an emergency. I one time needed to call my mother to ask who will pick me up from school that afternoon. Since I didn't have a phone at the moment, I had to walk to the office door, ring the buzzer, and wait for the office to unlock the door for me to access to the school provided phone. There usually is a small line and need to wait. If many students weren't allowed to use their devices in class, lots of students would need to interrupt their lesson would need to wait a huge line just to contact their parent or guardian. It would eventually lead to being a way for students to have an excuse to leave class. If students would have access to their devices in class, they could ask for permission to text whoever they need to contact at any time necessary.

Providing phones in the classrooms can be very useful in many different ways. To research, listen to music to help concentrate, and contact their parents or guardians when needed are three of many other positives that the phones can be as useful as an everyday school provided device. There may be some negatives, but there are also many positives of using these devices in classrooms that lots of school staff don't recognize. Phones can be helpful to the classroom environment and we should try to let the students use their phones in class to help make learning faster, easier, and a class friendly tool in various classrooms.

32

Student Name: _____

Prompt Number: 3

Writing Pilot Composition Page

The one Person that's important to me

I have that special someone in my life that can always count on. She could never be replaced for nothing in the world and that is my bestfriend Name ____.

My bestfriend and I have been through thick and thin. She is very important to me because she has never left my side for nothing in the world. We've always stuck together since fifth grade. We are like to peas in a pod. ____ is A person that will always have my back and that I can tell every thing too. I knew when I first met her that she was going to be the one that's going to be my side forever. We've only had one argreement and she still didn't leave my side. Sometime we get mad at each other but at the end of the day we are still bestfriends. She is really the only person that I can really talk to about certain things because I know that she can keep whatever I tell her. Every time she comes over or I go over house one of us has to tell something to one another.

My bestfriend is the main one I trust. She could never be replace no matter what I can always put my trust in her because I know that she will be the one to carry me through. I could never ask anyone else to be there for but her. She is like my blood sister to me. She always said to me when I'm feeling down that "I love you more than anything I'll never let any one come and take your spot. And that the truth.

A love my bestfriend She could never be traded for nothing that I really want. She will always be my bestfriend.

Dump Those Phones

Being able to bring phones to class has always been a major issue throughout many schools. However great these little machines may be, they are also the main cause of decreased focus and concentration in the classroom. Phones have proven many times to be a distraction to the student possessing the phone and those around them. Accordingly, phones should not be allowed in class because they draw the concentration of the class away from the discussion and can lead to decreased grades.

When students are given the privilege of having their phones in class, they eventually lose their concentration. Soon, they begin to pay more attention to the amount of time left in class then learning and comprehending the lesson. Although some may disagree and argue that teachers could take up the student's phones, the students still become anxious to play on their phones again, or text their family and friends. Also, phones in class have tended to change a student's priorities from studying and doing well in school to playing and texting on their phones. This has been proven by numerous researchers who report that 92% of students text with their phones during class. While these students are texting, they become distracted while they wait for a reply so they miss what the teacher is saying because they were focused on another subject.

In many schools, phones aren't allowed to be out during class. Due to this, many students attempt to sneak their phones, which is never an option or good thinking. While sneaking their phones, many people underestimate their teacher and eventually their phones are taken up. Once this happens, the students or their parents will have to pay in order to retrieve their phones. Of course some may say that the students could pay for themselves, but it is still the parents who provide them with money and it can be a

221

continuous burden for those who are tight on money. My friend   once ran into this conflict and became very overwhelmed.

She had used her phone during class to text her parents, but she had forgotten to ask so her phone was taken up. Her family was in a

very tight position so when the time came to pay the fine, her parents became very stressed and had to work extra. If phones weren't

allowed in class, then this wouldn't have occurred in the first place.

     As well as becoming a burden for most parents, cellphones in class can also become a serious distraction for the student who

owns the phone and for the students around them. This distraction can become an encumbrance for others when the owner has to

make an urgent phone call or text in the classroom. When a student has to text or call their parents during class, it becomes a major

distraction and can lead to loss of concentration and focus throughout the class. Even if students were to step out of the classroom and

make their phone call or text, it still becomes a distraction to those around when a person goes in and out of the room. Although there

are a list of positives, the list of negatives definitely overrules. Most of these reasons can become life-changing to the student's future

when they get to college, because these are the times that they need to concentrate the most. When phones are buzzing around the

classroom, you will become unaware of what your professor is telling you. Therefore, you miss points on the test because you didn't

have the correct notes. This occurred one day when my classmate sitting next to me got up to make a call. When she got up, I was

distracted by the sudden movement and my focus for the teacher was neglected and instead was drawn to my classmate. At that very

moment was when the teacher listed the most important facts on the review, and due to this, I missed those questions on the test.

Although phones are entertaining, they aren't always there for your benefit. They can be misleading and can lead you away from a successful future. So, why don't we dump those phones today?

Totem by Eamon Grennan

The poem Totem by Eamon Grennan was basically about a jack-o-lantern taking us through it's journey of going through death.

A pumpkin was put outside. Your probably asking, "Why was the pumpkin put out in the first place," and the answer is Halloween. In the beginning of the poem it said, "All Souls' over, the roast seeds eaten, I set on a backporch post our sculpted pumpkin under the weather, warm still for November." The line is just saying that "All Souls'" means November 2nd (Halloween), the point of view is from a pumpkin, the "backporch post" tells us the setting, "sculpted pumpkin" means that it's a carved pumpkin making it into a jack-o-lantern, "under the weather" is another way of saying feeling sick, and "warm still for November" means that as the sun beams down on it, you can conclude that the pumpkin is rotting faster by the minute. Another part of the poem is "it bows its boogeyman face of dread", the word 'boogeyman' is not a everyday poem word you would usually hear.

Overall the poem "Totem" was written to just tell us that eventually we will all reach the same fate, and we should just accept it and move on.

Thump! Thump! Thump! I heard footsteps dancing down the hall, creeping closer and closer like they were coming for me. Creeeek!! My whole body started to shake. The door opened. "   is up next for the district speech competition," the frightful voiced announced. I gulped down my feelings of fear and cautiously stepped onto the stage to see the wide-eyed audience staring back at me, waiting for me to speak. As I walked to the podium my feet suddenly felt like they were loaded with weights. Although I was trembling, I managed to take a deep breath and begin my speech. Words felt like they were creeping out of my mouth so slowly at first, but then I took another breath and felt myself relax, helping me to deliver the rest of my speech more confidently. I took one last glance at the audience before I left the stage. Were they smiling at me?

I felt so relieved and happy with my performance, but those feelings did not last as I sat and listened to the speeches of the other competitors. My heart sank more and more as each person delivered awesome speech after awesome speech. The fear and doubt about my speaking ability returned. I began arguing with myself, "You did so badly," the voice in my head said. "Be quiet," I tried to argue back. To make things words, the judges informed us that the results from the speech competition wouldn't be revealed until later that evening. UGH! The anxiety of not knowing how I did dragged on me as I got into the car. "It's going to be okay," I heard my mom's comforting voice say as I slipped on my seat belt. "The important thing is that you tried," my dad added. I know my parents were trying to help, but I still felt the exact same way I did before I even stepped on the stage.

Once we got back home, I sat down at my desk, put my phone to the side, and tried to do my homework. I tried and tried to do my homework, but my mind kept wandering back to my speech. I replayed it over and over I my head. "Did I speak clearly enough? Could everyone hear me? Did I speak with emotion and energy? Did I make good points?" I was driving myself nuts and was so thankful that mom interrupted my pity party and called me for dinner.

Just as we finished eating, the phone rang. My body went stiff, and I couldn't move a muscle...not even to reach for my phone. Dad reached over and picked up my phone. "Hello?" he answered. I heard nothing. Dad only nodded his head in response to the voice on the other end of the call. I held my breath, clenched my teeth, and closed my eyes. This one-minute call felt like pure torture. My dad finally hung up, stood face-to-face with me and took a slow, deep breath. "   I know you tried, so hard, but I have to let you know that YOU MADE IT TO THE REGIONALS!!!" At first my heart sank. Then, I realized what dad had just told me. I blinked my eyes, once, twice, and then screamed, "YAY!!" I almost burst into tears of happiness. Making it to regionals in speech was one of the most memorable events of my life and I will never forget it!

Page 8

225

Student Name:_

Prompt Number:____1____

Writing Pilot Composition Page

When I grow up I want to be a
lawyer because lawyers make alot of
money and lawyers also help people.
Lawyers help people with cases and
that's what I want to do, help people.
If I can't be a lawyer than I want
to be vetenerian because I can help cure
animals and I can see animals every
day all day. Vetenians and Lawyers are
totally different I don't know what is
the amount of money the vetenerens
make, but It looks like a good or fun
Job to work in, those two Jobs are the
only Jobs that I have interest in.
The only thing that I don't like about
being a lawyer is that It is a very
very long procces which takes extremely
long to take care of. And one thing
that I don't like about being
a vetenerian is that I have to deal
with animels that die and that's really
sad, or sometimes I will have to
deal with animels that need shots
and some dogs can be really egresstve when
It comes to shots.

Texas Writing Pilot 2016-2017-Draft

Page 5

226

APPENDIX J

Post Interview Protocol

Post-Interview Protocol

Thank you for taking the time to meet with me. The purpose of this final interview is for gathering a bit more information regarding the things I observed as you completed the think aloud protocols. It will also provide you an opportunity you to share any reflections you have as we look back on your experience of scoring student writing samples. As you answer each question, I may have a follow up question or two in order to gain a more complete picture of your initial answer.

In the first exercise you were not provided with a rubric, however in the second exercise you were provided with a rubric. Talk about your experience of evaluating student writing without and then with a rubric.

When you scored the student papers without a rubric you said _____. Talk to me a little bit about where you learned about that idea/concept.

When you scored the student papers with a rubric you said _____. Talk to me a little bit about where you learned about that idea/concept.

What other thoughts or reelections would you like to share about your experience in evaluating student writing samples.