Abstract: Standardized testing and accountability are currently unavoidable components of Texas Public Education. Through years of push-back, parents and educators have demanded that Texas consider alternative testing options that would reduce the high-stakes testing burden on students and schools. In 2015, the State of Texas passed legislation requiring the Texas Education Agency (TEA) to undertake a study of authentic writing assessment. This paper draws on data from a larger qualitative study to illustrate the complexity of teacher decision-making in the assessment process, provides further consideration into the influences of scoring calibration such as teacher knowledge, and highlights the need for intentionally designed professional learning about scoring as a means to mitigate differences and ultimately improve inter-rater reliability.

Keywords: assessment, calibration, inter-rater reliability, teacher decision-making, writing

Dr. Cato is an experienced educator and academic who likes to keep grounded in the work of practitioners in the classroom. Dr. Cato’s experience includes classroom teacher, curriculum director, and professional developer. She holds a PhD in Reading Education from Texas Woman’s University and an EC-12 Principal Certification, among others. Her research interests include the validity and reliability of assessment as well as the development of teacher knowledge and expertise. Her Twitter handle is @heathercato.

Dr. Walker worked as a K-12 Texas public school educator for seven years prior to moving into higher education. She served in the roles of middle-grades classroom teacher, middle-grades specialist, and K-12 instructional coach. Her research focus is on the intersections of adolescent literacy, ESL instruction, and middle schools. She also conducts research focusing on culturally responsive school leadership. Her Twitter handle is @k_walker6.
Working with teachers in the field of writing instruction and assessment is a familiar place for us. But on one particular fall day, when the lead researcher found herself before a room full of educators ready to learn how to score a new type of writing assessment, it didn’t feel like that familiar place to which she had grown accustomed. While her attendees arrived to the same training, in the same place and time, they brought with them a wide range of experiences, expertise, and knowledge when it came to writing assessment and instruction. She realized at that moment that the knowledge and experiences of the teachers in the room would color their understanding of the content and she began to wonder how she could leverage these different proficiencies to grow together in a common understanding of qualitative writing assessment.

On that day, the teachers in attendance 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 (Texas Education Agency, n.d.). 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 the lead researcher’s role with the project, she supervised the implementation of the pilot in the Dallas-Fort Worth area of the state. She served as a liaison between TEA and participating districts, as well as facilitated the assessment scoring training TEA created for the teachers.

The training she facilitated was created and given to her by the Student Assessment Division at TEA (n.d.). As she watched teachers interact throughout the training, she became increasingly concerned that the training was created with the assumption that all teachers arrived at the training with identical understandings of writing and writing assessment. As she continued to work with the teachers, she wondered what filters each were using to make sense of the information being presented in the training. Ultimately, these wonderings led her to an exploration of the complexities of teacher knowledge and experience within qualitative writing assessment and how these understandings influenced scoring decisions.

**Background on the Pilot**

With pressure mounting from parents and educators across the state to reduce the testing burden on students, the State of Texas passed legislation in 2015 (Tex. H.B. 1164, 2015) requiring the Texas Education Agency to conduct a study related to authentic writing assessments. A key feature of Texas Writing Pilot was to provide students with more timely feedback on their writing within the context of classroom instruction 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. Ideally, students would have access to feedback throughout the year to use to improve their writing in an on-going progression. In essence, the Texas Writing Pilot offered a testing format focused on growth of writing proficiency. The pilot made provisions for the students’ classroom teacher to assess student writing and provide timely feedback. Unfortunately, through a feasibility study conducted by TEA, which measured “the quality of locally-produced ratings” (Texas Education Agency, 2017), 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. The lead researcher 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.
Reliability in testing 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). In the TEA study, “the quality of locally-produced ratings” (Texas Education Agency, 2017), 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, 2015; 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, 2015; 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, 2015). To better understand why teachers, according to TEA did not produce consistent scores, factors such as a teacher’s experience and background must also be considered (Jeong, 2015).

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 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 a teacher goes throughout the school day she must carefully weigh the internal and external competing forces in her 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 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), yet teachers often struggle to explain how their beliefs influence their actions or decisions (Gill & Hoffman, 2009). Outside the four walls of their own classroom there are other factors 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-frame that they can 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 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 timely 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). 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 discipline they teach 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 opportunities, 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 they are 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 experiences 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 teaching practices (Penuel et al., 2007). When opportunities to engage and experience (Neuman & Cunningham, 2009) learning first-hand is followed by reflection, teachers 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 practices, they are able to develop shared understandings 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 practices (Borko, 2004).
Although teachers often 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 are 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 Kentucky Instructional Results Information System (KIRIS), a statewide standardized test consisting 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 considers the influences of teacher decision-making as well as explored the role professional learning plays in teacher decision-making.

Methodology

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 scoring student writing through the collection of multiple data types. The research question guiding this study was, “What sources of knowledge do teachers draw on when evaluating student writing?” In the following section, we 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 identified in The Texas Writing Pilot (TEA, n.d.), this study focused on participants who taught writing at a grade level for which writing is assessed by the state (i.e. fourth grade, seventh grade, and ninth grade). Through the use of purposeful sampling (Merriam, 1998), six participants within the Dallas-Fort Worth metroplex, one from each of the three grade levels,
(fourth, seventh, and ninth grades) were selected for the study. Of those six teachers, three had more than five years of teaching experience and are considered experienced teachers for the purpose of this study. The other three teachers had fewer than five years of teaching experience and are considered novice teachers for the purpose of this study. All participants had diverse personal and professional learning experiences that influenced their scoring decisions, such as years of teaching, where they taught, professional development on writing instruction, professional development on writing assessment, and personal experiences with writing. For this study, all names used are pseudonyms.

Methods of Data Collection

To better understand and consider the sources of knowledge teachers draw from when evaluating student writing, each participant completed 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, the lead researcher arranged a time to meet with each teacher face-to-face and engage with them in a semi-structured interview (Merriam, 1998). The interview data offered a richer description of their experiences in teaching writing as well as their experiences with professional learning related to writing instruction and assessment. See Appendix A for the survey and interview questions used.

During the time of the initial interview, the lead researcher asked the teacher participants to rate several student writing samples without using a rubric and using a think-aloud protocol (Ericsson & Simon, 1984; Smagorinsky, 1994). In this phase, the teachers vocalized their thoughts, explained what they noticed about the writing, and provided an explanation for how they might score the essay if they were to give it a grade. The aim was to gain insight about what the participants considered when making scoring decisions without being encumbered or guided by a rubric.

In the second phase, the participants were given the scoring rubric that was used for The Texas Writing Pilot and, again, used a think-aloud protocol to provide insight into their considerations when rating. The Texas Writing Pilot rubric was an unfamiliar tool to the participants, because all six of the participants had not previously participated in The Texas Writing Pilot. The participants took time to familiarize themselves with the rubric before scoring several more writing samples, this time using The Texas Writing Pilot rubric. As the participants rated the writing samples using the rubric, patterns began to emerge for the sources of knowledge teachers were drawing from when making scoring decisions.

After initial data analysis, the lead researcher 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). Participants reviewed the interview transcriptions and the think-aloud protocols and confirmed my interpretation of what they shared. No revisions to the transcripts were requested by the participants. In the final interview, participants were asked to review the transcription of them scoring the student writing samples. As the transcripts were reviewed, we would stop to discuss a particular idea or concept that they mentioned during scoring. For example, one teacher commented on 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 the lead researcher’s role as a liaison between TEA and the participating districts in North Texas, bracketing (Creswell, 2013) was necessary and maintained by ensuring that the lead researcher kept solely within the scope of the pre-established protocols. In the
following section, methods are explained to ensure the trustworthiness (Merriam, 2001) of the study.

**Researcher Subjectivity**

The authors of this paper are white, middle-class, female, middle-grades literacy educators from Texas. The authors attended universities in West and North Texas and taught in K-12 public schools in West, North, and South Texas. Additionally, both authors have served in central office administrative positions and in higher education. We believe that educators are competent professionals and autonomous actors that are tasked with meeting the diverse needs of the students in their classrooms to produce growth in learning and skill development. It is our belief that literacy education should be treated as an on-going progression of skills and concept knowledge that are explored and enhanced through authentic writing opportunities and text-embedded student discussions. Thus, we believe that providing teachers with high-quality professional development opportunities to increase their knowledge and skill in instruction and assessment is key to maintaining dignity within the profession and improved student learning outcomes.

**Data Analysis**

The aim of this study is not to generalize the findings to a given population, but to consider the framework of scorer cognition established by Wolfe & Feltovich (1994) to inform future work related to alternative assessment design for the purposes of high-stakes testing. 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, we came across the research study of Wolfe & Feltovich (1994) who considered the cognition of the scorer during the evaluation process. In their study, Wolfe & Feltovich collected data from novice and expert scorers as the participant scored student writing for a large-scale standardized assessment using a 6-point holistic rubric. The research by Wolfe & 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 & Calfee, 1983) who considered the process by which scorers processed information as they evaluate writing. Scoring writing is a complex decision-making process that calls for the scorer to draw upon several interpretive frameworks to make a scoring decision.

In order to investigate the knowledge teachers draw upon when evaluating student writing, the primary unit of analysis was each teacher’s verbal statements to each of the questions asked and each statement they made during the think-alouds. Statements from the participants revealed their processes of pedagogical reasoning and action (Shulman, 1987). To begin the analysis process, we coded the data from both the teacher’s think-alouds through the lens an interpretive framework Wolfe & Feltovich (1994) identified as Source Knowledge. According to Wolfe & Feltovich, Source Knowledge is a comparison processing action “performed by manipulating some external form of knowledge.” (Wolfe & Feltovich, 1994, p. 36). In their study, Wolfe & Feltovich, 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). When coding the data, we reviewed each verbal statement by the participant to determine which of the three Sources of Knowledge identified most appropriately matched the statement.

Wolfe & Feltovich’s (1994) Sources of Knowledge provided a framework for understanding how personal and professional experiences influence interpretations of student writing. However, it became apparent that there were some statements
that did not easily fit into one of the three original codes. To better categorize each statement made by the teachers, we chose to broaden the definitions of the three mediums (Wolfe & 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, the following categories were added: 4) Assignment (knowledge from experience with writing assignments). The use of open- and closed-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 into the knowledge teachers use to make scoring decisions by allowing new sub-categories to emerge. See Table 1 for Source Knowledge definitions.

Using the transcribed responses from the think-aloud-protocols, eleven different sub-categories of Source of Knowledge were identified using teacher comments. Appendix B shows the categories and sub-categories identified by coding the think-aloud protocols and includes an example of each sub-category from the data.

As an additional layer of the data analysis, the transcripts from the post-scoring interview were coded, where teachers self-identified their Source Knowledge from statements they made as they scored student writing, using the eleven sub-categories of Source Knowledge identified in the initial phases of coding. This layer of coding confirmed the eleven sub-categories that we had identified in the initial coding. The Personal Experience sub-category was further explored, and nine more narrowly defined sources of personal experience identified by the teachers. Table 2 provides a list of those nine sources.

It was through breaking the data down into these more narrowly defined categories that patterns within the data became clearer and began to emerge from the larger category of Prior Knowledge. It became clear 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 we will provide a narrative of each participant and further analysis of these three influences.

---

Table 1

Source Knowledge Definitions

<table>
<thead>
<tr>
<th>Source Knowledge Code</th>
<th>Wolfe &amp; Feltovich (1994) Definition</th>
<th>New Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>other papers previously read by the scorer</td>
<td>any prior knowledge considered in the scoring process</td>
</tr>
<tr>
<td>Scorer</td>
<td>other scores assigned by other scorers</td>
<td>knowledge from experience scoring writing</td>
</tr>
<tr>
<td>Rubric</td>
<td>descriptions provided in the rubric</td>
<td>knowledge from experience using rubrics for scoring</td>
</tr>
</tbody>
</table>
Findings

In what follows, we 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 experiences 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 twenty years of experience teaching, Debbie taught 6th - 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 twenty 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 Debbie explained that learning how to teach writing was more about:

... 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 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 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 isn’t 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 can 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 powerful learning experiences for educators looking to improve their

<table>
<thead>
<tr>
<th>Types of Personal Experience</th>
<th>Experiences with Writing</th>
<th>Professional Learning Experiences</th>
<th>Peers and Mentors</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience as a student in college or high school</td>
<td>Structured learning experiences</td>
<td>Peers and Mentor relationships</td>
<td>Unknown/ not sure of source</td>
<td></td>
</tr>
<tr>
<td>On the job experience as a teacher or in another profession</td>
<td>Self study and reading experiences</td>
<td>Team calibration experiences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>District lead professional learning</td>
<td>Team/Committee curriculum writing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
craft. This section will explore learning experiences that teachers drew from when making writing scoring decisions.

**Donna, an experienced seventh-grade teacher**

Although for the last twenty-four years, Donna has taught middle school students reading and writing in the suburbs of Dallas Fort Worth, her first career was as 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 the lead researcher 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 the lead researcher 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 five 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 three-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, “It was teachers teaching teachers and they did a good job of showing us ‘here’s what to do,’ ‘here’s why it’s important,’ and ‘here’s why it works.’ 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 the National Writing Project it is the intention of both organizations to get teachers writing themselves as well as learning from other teachers. For both 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**

Teaching is a never-ending process of learning and improving, not only for students, but for teachers as well. While many learn from past and professional learning experiences, 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 we will 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 twenty 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 “the whole time you are there you’re a student, so everything you do is like you’re 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 four 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 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 mentioned for Sarah, her mentors came 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. Vera John-Steiner (1985) termed this experience "the legacy of their distant teachers" and explained that this distant 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 illustrated 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 suggested
that personal experience, professional learning, and mentorship influence teacher decision-making. This study provided 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 our experience through the interview process that teachers often adequately articulated 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, 2015; Lumley, 2002). Ultimately, what an educator believes about teaching and learning, can in fact, directly influence her decisions (Gill & Hoffman, 2009; McMillan, 2005; Putnam & Borko, 1997).

Teachers found a level of meaning and usefulness from the professional experiences that they mentioned...they felt empowered to turn and use the newly acquired knowledge in their classroom.”

From the data in this study, we concluded that teachers most often draw from their experiences that include personal experience, professional learning and inquiry, and 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, 2015; Knoch et al., 2007).

In contrast to the findings in this study, the findings from the study conducted by TEA indicated that teachers did not score reliably consistent enough for this type of assessment to scaled statewide (Texas Education Agency, 2017), the reports also indicated 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 off one day professional development. This suggests that the quality inter-rater reliability was directly related to the type of professional development or training raters received prior to engaging in the assessment process.

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 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 TEA had assigned each paper. This type of mechanical 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, 2015; 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.

The findings of this study, when compared to the findings of the TEA (Texas Education Agency, 2017), point to a failure in the training process rather than a failure in assessment design or the abilities of teachers to engage in high-quality qualitative writing assessment while maintaining strong inter-rater reliability. The assumption from the TEA (Texas Education Agency, 2017) study that teachers were not capable of participating in reliable qualitative writing assessment led to assessment and accountability decisions in writing that directly impacted instruction and learning in the classroom, as the State used the data to justify the continuation of high-stakes writing assessment (Texas Education Agency, 2022).

Schools are currently caught in the middle of a culture war and the result of high-stakes assessment when paired with legislation regarding Critical Race Theory and banned books has robbed educators of professional autonomy, limited local control of schools, and ultimately has left students at all grade levels less college ready (Moody, 2022). The authors of this paper call for a re-examination of writing instruction and assessment at the State-level. Based on the findings of this study, we suggest that high-quality professional development for teachers in writing assessment and writing assessments that focus on student growth through the writing process and development will lead to an improved inter-rater reliability on writing assessments and writing instruction that will promote college-readiness for students (Graham, 2019).
References


Appendix A

Participant Questions

Participant Questionnaire

- Name (first and last)
- I have previously participated in the Texas Writing Pilot
- How many years have you taught writing?
- What grade levels have you taught?
- What teaching certifications do you have?
- How would you describe yourself as a writing teacher?
- How would you describe yourself as a writer?
- As it relates to writing instruction, in what experiences or specialized training have participated (e.g. college courses, workshops, conferences, coaching)?
- Describe any experience you have had with teaching or preparing students for a state writing assessment?
- Would you be willing to further participate in the research study during the spring semester 2019 by taking part in interviews and completing a writing scoring protocol?
- Phone number (best number to reach you)
- Email address

Participant Pre-Scoring Interview

- As it relates to writing instruction, what experiences or specialized training have you had?
- Have you ever taught or prepared students for the state writing assessment? If yes, what training have you had to prepare students for the state writing assessment?
- How long have you been teaching or preparing students for STAAR, or how many years have you been teaching students for state assessment.
- Do you consider yourself a writer?
- What would you say are your beliefs about writing, and not just in school, just in general?
- Please describe how writing instruction typically looks in your classroom?
- How did you learn to teach writing?
- Please describe the ways in which you currently assess student writing.
- You mentioned training that you have attended, what are some takeaways you had from the training and how that was utilized in your classroom?

Participant Post-Scoring Interview

- 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 with and without a rubric.
- What kinds of training have you had on the use of rubrics?
- What kinds of things did you do during those training sessions?
- How do your experiences with using rubrics influence how you score?
- When looking at multiple expectations within a single score, how do you determine between the two scores?
- What if I told you this rubric was used in fourth, seventh and ninth grade the same exact rubric, the same exact words?
Is there anything else about the use of rubrics you would like to share?

Appendix B

Sources of Knowledge

<table>
<thead>
<tr>
<th>Categories</th>
<th>Sub-categories</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge</td>
<td>Teacher’s Personal Experience</td>
<td>“That’s definitely a pet peeve of mine that I noticed especially as a journalism major.”</td>
</tr>
<tr>
<td></td>
<td>Student Ability - related to the age of the student</td>
<td>“If this was from a ninth-grade class, it would get a low grade.”</td>
</tr>
<tr>
<td></td>
<td>Student Ability - related to English Language Learners</td>
<td>“I think this is almost an EL student, like an English Language Learner.”</td>
</tr>
<tr>
<td></td>
<td>Student Ability - related to Special Education Learners</td>
<td>“...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.”</td>
</tr>
<tr>
<td></td>
<td>Writing Instruction/Writing Technique</td>
<td>“So she’s probably being taught how to do [that kind of] description.”</td>
</tr>
<tr>
<td></td>
<td>Personal expectations for writing or writing instruction</td>
<td>“What am I looking for like some suspense...”</td>
</tr>
<tr>
<td>Assessment Knowledge</td>
<td>Relationship or connections to the directions of the assessment</td>
<td>“I’m wondering if there was an amount of words they had to use?”</td>
</tr>
<tr>
<td>Scorer Knowledge</td>
<td>Relationship or connections to other student writing papers</td>
<td>“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...”</td>
</tr>
<tr>
<td></td>
<td>Relationship or connections to other assessments</td>
<td>“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.”</td>
</tr>
<tr>
<td>Rubric Knowledge</td>
<td>Relationship or connections to the rubric expectations from the assignment’s rubric</td>
<td>“This one right away I know is different from the other one this is probably not gonna get any sixes or fives.”</td>
</tr>
</tbody>
</table>
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...”