Lessons Learned From the Texas Writing Pilot

By Heather Cato

Heather Cato is an independent educational strategist and a teaching associate at Texas Woman's University. Her research interests include writing assessment, teacher decision making, and professional learning. She can be reached at heatheracato@gmail.com or on Twitter @heathercato.

Abstract: In 2016, with mounting pressure from parents, educators, and other stakeholders around the state pushing back against the volume of standardized tests students are required to take, the Texas Education Agency introduced the Texas Writing Pilot. This pilot was designed to study the feasibility of creating an alternative writing assessment that would assess a student's mastery of the essential knowledge and skills in writing through a portfolio writing assessment. Excitement for a different type of assessment rose across the state as many districts eagerly applied to participate in the pilot. But after three years of work from educators across the state on the pilot, the Texas Education Agency abruptly ended the project.

Findings from this document analysis will bring 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 needed to more effectively teach writing.

Keywords: state assessment, authentic assessment, portfolios, writing, advocacy

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 means through which students demonstrate their writing skills. So, in 2016, when I had 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, three 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 Texas Education Agency (TEA) and participating school districts. 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. But after three years of work on the pilot, TEA 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 TEA 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?

The answer to that question starts with a bit of background on the origins of the Texas Writing Pilot. In 2015, there continued to be an outcry from educators and parents across the state to reduce the burden of standardized assessments on our students. In response, Representative Gary VanDeaver introduced Texas HB 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. (HB 1164, 2015)

In response to the legislation, the Student Assessment Division at TEA worked with the Educational Testing Service (ETS), the company that held the state's assessment contract, and representatives from Representative VanDeaver's office to design the parameters for the Texas Writing Pilot (Morgan, 2018). 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 (Texas Education Agency, 2017). The timeline for submitting the four writing samples was spread throughout the course of the school year. Figure 1 depicts the TEA Timeline for the Texas Writing Pilot.

The committee I was initially part of advocated that students needed 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. To provide this timely feedback to the students, the design of the pilot included that the student's classroom teacher would assess the student writing as opposed to sending it off for a blind scorer to rate (Texas Education Agency, 2017). Ideally, students would be able to use this specific feedback as they continued to develop their writing skills throughout the year.

Other parameters of the pilot stipulated that these additional writing samples should mirror classroom writing instruction and incorporate the writing process from start to finish (Texas Education Agency, n.d.). 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 (Texas Education Agency, 2017). 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.

Many administrators and educators praised the pilot and reported students were writing more and the quality of writing instruction was better because it was unencumbered by the strain of the State of Texas Assessments of Academic Readiness (STAAR) test (Morgan, 2018; Texas Education Agency, 2018). However, in the final report that 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).

With so much promise, it was disappointing to read TEA's conclusion that the Texas Writing Pilot was not a valid assessment. Although I had been quite involved with the pilot from the beginning, I felt it necessary to approach the evidence with fresh eyes. In the following section, the theoretical framework sets the groundwork to explain the lens that I used to approach this study.

Theoretical Framework

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). In order to build the knowledge structures (Harel & Papert, 1991; Papert, 2006) necessary for this research, I chose to utilize the lens of constructivism. 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" (Picard et al., 2004) as they reflect upon the learning and internalize the new
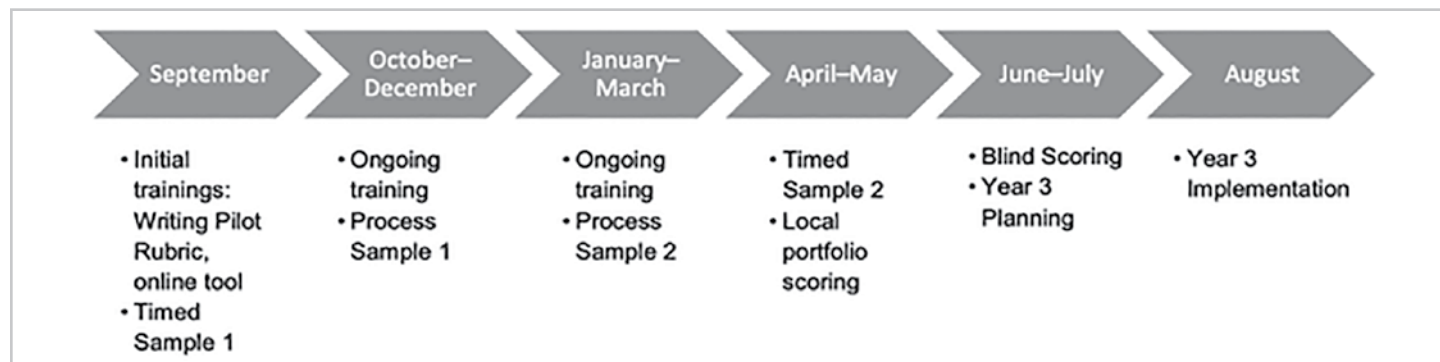
Figure 1. TEA Timeline for The Texas Writing Pilot

knowledge. Because of my dual role as a participant in the Texas Writing Pilot and as a researcher looking back upon what occurred during the pilot, I relied on constructivism as the means for which to review each of the documents.

Literature Review

In reviewing the documents related to the Texas Writing Pilot, this document analysis seeks to explore the question of what happened during the Texas Writing Pilot. Because the state of Texas ended the pilot citing "scoring correlations and rater-agreement never reached the same level as STAAR, at scale" (Texas Education Agency, 2018, p. 2), I felt it necessary to focus the literature review on previous research that gave further insight into the reliability and validity assessment.

Assessment Reliability

As far as testing is concerned, reliability can be defined as the frequency with 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, 2015). 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).

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 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).

While rater agreement is an important consideration in measuring the quality of an assessment, the research suggests that reliability and validity are more complex that merely if a couple of raters came to any kind of consensus on a score for a given paper.

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 HB 1164 (2015) as it pertains to student performance assessments. I collected data 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 HB 1164 (2015), the original press release and request for participants, and the three published program reports submitted to the Governor and the Texas Legislature.

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 (Merriam & Tisdell, 2016) 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 the key theme(s) or idea(s) of the line was and then I recorded my findings in a spreadsheet where I collected the findings of each of the reviewed documents. This analysis provided the opportunity for categories and themes to emerge and included careful reading and re-reading of the data (Bowen, 2009). During the analysis of each document, I engaged in member checking with other school leaders who had taken part in the design and implementation of the pilot to ensure the accuracy of the coding (Merriam & Tisdell, 2016).

Findings and Discussion

Data analysis led to the initial development of 29 identified codes. From additional analysis of these codes, six categories emerged: Authentic Assessment, Writing Portfolios, Student Growth, Instruction, Scoring: Rater Agreement, and Training. Table 1 provides a look into each category and the codes that were associated with each category, along with excerpts of supportive data revealing insight into what happened during the pilot.

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. The following sections provide an explanation and elaboration for each of the six categories.

Assessment Design

The language of Texas HB 1164 (2015) was clear that the design of the writing assessment should be authentic in demonstrating student growth in writing skills across multiple genre. So, it is no surprise that much attention in the Texas Writing Pilot documents was given to the actual design of the assessment. Upon careful review of the documents, it became evident that there were three specific areas of focus for the design of the assessment. These three areas—authentic assessment, writing portfolios, and student growth—became the headings for three of the categories which I will discuss here.

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).

Table 1. TEA Timeline for The Texas Writing Pilot

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. During an initial face-to-face meeting in 2017, participants collaborated to develop the foundation of the writing pilot rubric (Texas Education Agency, 2017). This rubric was then used to give each student writing sample a holistic score. However, the documents revealed that Year 1 participants viewed the rubric too similar to the STAAR rubric (Texas Education Agency, 2018) and pushed for a new rubric to be created and used in Year 2. This new rubric shifted the pilot rubric from a 4-point holistic rubric to a 6- or 3-point analytic rubric (Texas Education Agency, 2018).

I found statements that described the importance of the writing process within the creation of the students' writing portfolio (Texas Education Agency, n.d.) which were included in the writing portfolios category. However, there was a difference between the initial design and the actual implementation. The pilot design stated there would be three process writing papers (Texas Education Agency, 2016b), while the actual pilot sampled and assessed only two (Texas Education Agency, 2017). Although 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 (Texas Education Agency, 2017). TEA reported that this design feature allowed teachers to "fully align the assessment with local instruction and scope and sequence of curriculum" (Texas Education Agency, 2018, p. 7). However, district administrators complained that the timelines for the pilot did not correlate with other district demands, curriculum, and responsibilities (Morgan, 2018).

With the goal being to assess student mastery from the beginning to end of the school year (HB 1164, 2015), student growth is also an important category that was discovered from the documents. 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). Even teachers reported seeing student growth as they observed students showing stronger engagement with their writing assignments as a result of the pilot (Texas Education Agency, 2018).

Because interrater reliability never reached the same level of STAAR, TEA determined the design of the pilot was not something that would be able to be scaled statewide for the purpose of accountability (Texas Education Agency, 2018). However, in reviewing the data, it was clear that the design of the pilot drastically changed from Year 1 to Year 2. In the two years of implementation of the pilot, two different rubrics were used, the submission timeline was inconsistent, the choices of prompts for the timed writing were reduced, and participation numbers drastically increased from 1,700 in the first year to over 30,000 students in the second year (Texas Education Agency, 2018). Even though teachers reported that the assessment was authentic and more congruent with classroom instruction, TEA found the consistency in scoring to be lacking (Texas Education Agency, 2018), but any one of these variables alone could dramatically impact the results of interrater reliability. While supporters of authentic assessment should continue to advocate for other studies and opportunities that support core instruction through the use of authentic assessments, advocacy must also focus on the consistent implementation within the study.

Instruction

For the category of instruction, teacher reflections revealed that educators felt writing instruction was more intentional and focused because of the pilot (Texas Education Agency, 2017). While the study conducted by TEA did not evaluate alignment between instruction and assessment, teachers reported stronger congruence (Texas Education Agency, 2018). Teachers felt that even the prompts were more authentic because the writing prompts were a direct extension of what they were already doing in class (Texas Education Agency, 2017).

The intention of the pilot was that the writing assessments would be an extension of everyday classroom instruction (Texas Education Agency, n.d.). Throughout the documents, it was clear that throughout the pilot, teachers observed improvement in the quality of their writing instruction (Texas Education Agency, 2017; Texas Education Agency 2018). Teachers were also able to see a tighter alignment between what is taught and what is assessed because the assessment was integrated with 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 (Texas Education Agency, n.d.). Teachers reported that this reprieve from the state assessment provided them more leeway in teaching (Texas Education Agency 2017). As a result, teachers felt the quality of instruction was improved because they were not having to tailor their instruction to one specific test or style of writing (Texas Education Agency 2017). 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 (Texas Education Agency, 2018). Teachers expressed concerns that the way they needed to teach students for the pilot versus the constrained 26 lines for STAAR made them feel that there were competing forces at play (Morgan, 2018). 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: Rater Agreement

One of the main arguments TEA cited for discontinuing the pilot was that the pilot could 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. 2). Overall, there were four writing assignments for raters to agree upon. Two of the four pieces of writing students turned in were timed writing samples where students had to select from a list of previously released STAAR expository prompts (Texas Education Agency, 2017). This provided a relatively controlled field of responses for the graders but did not match the types of writing happening in the classroom (Morgan, 2018). The other two pieces the students wrote were the process pieces where either the teacher or the student chose the topic within a given genre for the student to write about (Texas Education Agency, 2016b). 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 (Texas Education Agency, 2018).

TEA acknowledged that the lack of rater agreement may have been due in part to "limited appropriations to the project reduced the ability for true piloting of a standardized assessment prototype"

(Texas Education Agency, 2018, p. 2)—meaning there was a lack of funding for training and calibration. While the sample scoring papers provided by TEA 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. 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 to

be utilized. There is a clear need here for TEA to provide further support as how a specific rubric should be applied. 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 they view as clear and distinct markers of quality writing.

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 and 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). Those of us who were involved in the pilot were witness to the insufficient time and support for teachers to calibrate. Another issue was the timing of the training. In Year 1, the scoring training did not occur until after the student writing was collected (Texas Education Agency, 2017), leaving teachers unclear as to the expectations of the writing assessment. In Year 2, with the increase in pilot participation not occurring until the second semester (Texas Education Agency, 2018), most teachers did not receive training for the writing samples until after three of the four writing samples had been collected.

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. There is an opportunity here for teachers to advocate for time to work together in calibration meetings in order for meaningful discourse about writing to take place.

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 about 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 led 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 seek new and alternative ways to authentically assess what a student knows, educators and administrators must take an active role by 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 had the opportunity to specifically influence 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 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 more than 28,000 students to a study in the last five months of the study burdened the entire study and calls into question its validity. The study did, in fact, provide both quantitative and qualitative data regarding the 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 for assessment purposes so that 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, TEA needs 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 HB 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 that when given the opportunity to attend a meeting, say yes— even when you are not 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 through 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.

References

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

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27.

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). National Council of Teachers of English.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE Publications.

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

Gibson, W., & Brown, A. (2009). *Working with qualitative data*. Sage.

Harel, I., & Papert, S. (1991). *Situating constructionism*. Ablex Publishing Corporation.

HB 1164, 84th Leg. R.S. (Tex. 2015).

Huot, B. (2002). *Rearticulating writing assessment for teaching and learning*. 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. (2015). What is your teacher rubric? extracting teachers' assessment constructs. *Practical Assessment, Research & Evaluation*, *20*(6), 1-13.

Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. Farrar, Straus and Giroux.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation*. Jossey-Bass.

Morgan, C. (2018). Educator perceptions of forces influencing implementation of a statewide writing portfolio assessment [Doctoral dissertation, University of North Texas]. ProQuest Dissertations & Theses Global.

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

Papert, S. (2006). Minding change. *Human Development*, *49*(4), 239-247. https://doi.org/10.1159/000094373

Picard, R. W., Papert, S., Bender, W., & Blumberg, B. (2004). Affective learning—a manifesto. *BT Technology Journal*, *22*(4), 253.

Texas Education Agency. (2016a). Request for applications: The 2016-2018 writing pilot program. 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. 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. 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. https://tea.texas.gov/student-assessment/testing/staar/the-texas-writing-pilot-program

Texas Education Agency. (n.d.). The Texas writing pilot program. https://tea.texas.gov/student-assessment/testing/staar/the-texas-writing-pilot-program

Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan*, *75*(3), 200-208.

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