2022

Examining The Quality Of Teacher-Selected Social Studies Performance-Based Assessments: A Responsive Evaluation

Lynne Mardigian Bland
William & Mary - School of Education, Lynne_Bland@ccpsnet.net

EXAMINING THE QUALITY OF TEACHER-SELECTED SOCIAL STUDIES PERFORMANCE-BASED ASSESSMENTS: A RESPONSIVE EVALUATION

A Dissertation

Presented to the

The Faculty of the School of Education

The College of William and Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

By

Lynne Mardigian Bland

EXAMINING THE QUALITY OF TEACHER-SELECTED SOCIAL STUDIES PERFORMANCE-BASED ASSESSMENTS:

A RESPONSIVE EVALUATION

By

Lynne Mardigian Bland

Approved December 2021 by

Megan Tschannen-Moran, Ph.D.
Committee Member

Leslie Grant, Ph.D.
Committee Member

Christopher Gareis, Ed.D.
Chairperson of Doctoral Committee

Dedication

For my husband, Jamie, and my daughter, Lizzie...you gave me love, purpose, strength, motivation, and unconditional support during my very long journey.

I am extraordinarily grateful.

Acknowledgments

I would like to extend my sincere thanks to my dissertation committee chair, Dr. Christopher Gareis, for his continued patience and support during my coursework and doctoral study. I would also like to thank my committee members, Dr. Leslie Grant and Dr. Megan Tschannen-Moran for their support. I would like to acknowledge Sara, who encouraged me to follow my educational path to the College of William and Mary, as well as Tricia for her continuous encouragement. I am deeply grateful to my father and his words of wisdom and confidence and for reading over many of my drafts, providing insightful comments and helpful suggestions. I thank my sister Kate who was one of my biggest cheerleaders and always believed in me, especially when my confidence would wane. I am deeply grateful to my friend and colleague Laurie who was a calming presence through my doctoral journey, was always willing to listen, provide feedback, and pull me into prayer when I needed reminding that all is in God's hands. I am grateful for my friend Ann and our "cohort of two," as well as Bettina and Jess for helping me flesh out my ideas and always telling me, "You've got this, Lynne." I am so appreciative for many other family and friends who cheered me on. I am also thankful for my dog Ling Ling who kept me company while I studied. I would finally like to thank my mother for her unwavering support and belief in me over the course of my entire life. She knew that I could do whatever I set my mind to. Though she will not be here to see me graduate, I know her spirit will be. I miss you and love you, Mom.

TABLE OF CONTENTS

List of Tables

List of Figures

Abstract

In 2014, the Virginia Department of Education (VDOE) communicated a plan to replace five Standards of Learning (SOL) assessments with alternative assessments based on legislation passed by the Virginia General Assembly (VDOE, 2014b, p. 1). School districts were able to choose the type of alternative assessment to implement, yet they were encouraged to use authentic performance assessments and a portfolio approach. One suburban school district in central Virginia chose a portfolio approach as the alternative assessment. Drawing upon the transdisciplinary approach of program evaluation (Mertens & Wilson, 2012), this study examined the quality of teacher-selected performance-based assessments implemented with students as one artifact within a collection of assessments available in the portfolio in two middle school social studies courses. The *Virginia Quality Criteria Review Tool for Performance Assessments* was employed as the instrument to examine the assessments for quality. The aim of this study was to strengthen the practice of consistently implementing quality performance assessments with students by answering the evaluation questions: To what degree do performance assessments used as local alternative assessments to Virginia Standards of Learning tests for accountability meet the state's criteria for quality? How are performance-based assessments being implemented in United States History to 1865 (USI) and United States History 1865 to the Present (USII) courses in PCPS? What are USI and USII teachers' perspectives on the merit and worth of the performance assessment process in PCPS? What are USI and USII teachers' recommendations for developing and using quality performance assessments in PCPS moving forward?

EXAMINING THE QUALITY OF TEACHER-SELECTED SOCIAL STUDIES PERFORMANCE-BASED ASSESSMENTS: A RESPONSIVE EVALUATION

CHAPTER 1

INTRODUCTION

During the past two decades, K-12 education assessment practices have favored objectively scored type questions, such as multiple-choice, over rubric scored performance-based type questions. There is some evidence to suggest that the K-12 federal law for general education, No Child Left Behind Act of 2001 (NCLB), may be the reason for this shift (Brookhart, 2013; Stecher, 2010). The law required states to report to the federal government the percentage of students who were proficient in reading and mathematics and to disaggregate the data by subgroups such as poverty, minority status, English proficiency, and disabilities (Brookhart, 2013). Objectively scored assessments were perceived as a quick and efficient way to report the high-stakes data in a timely manner. Schools not making adequate yearly progress 2 years in a row received rigid sanctions. "As a result of the NCLB law, unintended consequences arose: curricular content was narrowed to tested subjects, subject area knowledge was fragmented into test-related pieces, and teachers increased their use of teacher-centered pedagogies" (Au, 2007, p. 258). Narrowing the scope of curriculum and instruction promoted the practice of teaching to the test (Au, 2007; Brookhart, 2013).

Statement of the Problem

High-stakes tests eventually received backlash due to the measurement of minimum competencies (Brookhart, 2013). The counter argument, however, was that

> challenging tasks are more difficult to write than basic skills items, take time for students to do and for accurate measurement students need to complete many different tasks. These tasks require human scoring, which in turn requires time, training and money. Efficiency and cost ultimately won out. (Brookhart, 2013, p. 63)

Subsequently, opportunities for teachers to engage in professional development geared toward recognizing, creating, and implementing quality performance assessments diminished, as well as opportunities for students to engage in deeper learning. Therefore, the problem addressed in this study was the potential lack of quality performance assessments available to students. To specifically address this problem of practice, the current study examined performance-based assessments identified as one of the student artifacts within a portfolio approach to assessment for quality characteristics using the Virginia Quality Criteria Review Tool for Performance Assessments in a suburban Virginia school district.

Statement of Context

Federal Context

In time, the federal K-12 education law was revised to its most current version and became known as the Every Student Succeeds Act (ESSA), signed by President Obama on December 10, 2015. ESSA gives local school divisions more autonomy to improve student performance and gives educators a greater sense of ownership over school improvement efforts (Klein, 2016). Under NCLB, districts had to choose from a list of federal improvement options. Under ESSA, districts and even schools have the ability to create their own improvement strategies, as long as there is evidence to back up their approaches (Klein, 2016). John B. King Jr., former United States Secretary of Education, considered ESSA a “game changer because it gives districts and schools room to maneuver, while still focusing on strategies that have the

potential to yield student-achievement dividends" (Klein, 2016, p. 4). Under ESSA, states were required to develop new federal accountability systems during the 2016-17 year, to be implemented during the 2018-2019 school year (ESSA, 2017). Former Virginia State Superintendent of Schools, Dr. Steven Staples, wrote in a July 27, 2016, letter to the United States Department of Education that "Virginia applauded the bipartisan measure that reauthorized the *No Child Left Behind Act of 2001* and that the law's intent to restore autonomy to states over key accountability tenets provides welcome relief from the prescriptive requirements of NCLB" (ESSA Communications, 2017).

Staples went on to state that ESSA

> comes at an ideal time in Virginia, as our state Board of Education develops a new vision for the commonwealth's public schools that focuses on continuous improvement for all schools and on aligning student outcomes with the expectations of higher education and employers from all sectors. (ESSA Communications, 2017)

State Context

Prior to Dr. Staples's letter to the United States Department of Education, the Virginia General Assembly had already responded to increased concern regarding the amount of testing in local school divisions and the time spent in test preparation activities (Virginia Department of Education [VDOE], 2014). This resulted in a legislative mandate in the forms of House Bill 930 and Senate Bill 306. This mandate amended the *Code of Virginia* in 2014 to eliminate several SOL tests. *Code §22.1-253.13:3* states (VDOE, 2014, p. 1, emphasis in original):

> The Standards of Learning assessments administered to students in grades three through eight **shall not exceed** (a) reading and mathematics in grades three and four; (b) reading, mathematics, and science in grade five; (c) reading and mathematics in grades six and

> seven; (d) reading, writing, mathematics, and science in grade eight; and (e) Virginia Studies and Civics and Economics once each at the grade levels deemed appropriate by each local school board.

The code further required that schools document their instruction and alternative assessments as follows:

> Each school board shall annually certify that it has provided instruction and administered an alternative assessment, consistent with Board guidelines, to students in grades three through eight in each Standards of Learning subject area in which a Standards of Learning assessment was not administered during the school year. Such guidelines shall (1) incorporate options for age-appropriate, authentic performance assessments and portfolios with rubrics and other methodologies designed to ensure that students are making adequate academic progress in the subject area and that the Standards of Learning content is being taught; (2) permit and encourage integrated assessments that include multiple subject areas; and (3) emphasize collaboration between teachers to administer and substantiate the assessments and the professional development of teachers to enable them to make the best use of alternative assessments.

With the elimination of some tests used for accountability, the legislation's intent was to encourage greater use of *assessments designed to inform instruction.* The legislation did not mandate the exact type of assessment to be administered but hoped such assessments might require students to perform a task or create a product, typically scored by a rubric, that might occur in a real-life situation and provide information about what students had learned as well as the concepts and skills they had not yet mastered (VDOE, 2014, p. 2). A flexible 5-year timeline was put into place and was meant to

> provide an opportunity to engage in innovation to provide new opportunities for students to demonstrate their knowledge of the curriculum while simultaneously reassuring content teachers in later grades that the content upon which their own instruction was dependent had been taught, and that students came to them prepared to learn. (VDOE, 2014b, p. 2)

Expectations for including the SOL in alternative assessments were made clear while continuing to maintain flexibility. For the 2014-15 school year, alternative assessments were to incorporate each strand or reporting category for that content area and grade level; however, the assessments were not expected to cover all of the content standards contained in that strand or reporting category (VDOE, 2014b). The guidelines stated that in subsequent years, requirements for content coverage would be further clarified based upon the experiences in implementing the local assessments statewide that were gained (VDOE, 2014b). There was no requirement for local alternative assessment scores to be reported to the Virginia Department of Education for accreditation purposes (VDOE, 2014b). "Local school boards and division superintendents certified through the annual Standards of Quality (SOQ) compliance assurance that local alternative assessments measured the SOL and adhered to the Board's guidelines" (VDOE, 2014b, p. 2). For the first year of implementation, school divisions were asked to retain documentation of the alternative assessment requirements, copies of assessments, and ancillary materials such as rubrics or sample student work. This documentation would be available for desk reviews by VDOE staff as evidence of implementation and as student exemplars to strengthen alternative assessments statewide (VDOE, 2014b). In addition, school divisions were asked to prepare plans that described how local assessments were designed to inform instruction and how they would be expanded in subsequent years (VDOE, 2014b). The legislation

encouraged integrated assessments that included multiple subject areas while being aware of the Individuals with Disabilities Education Act (IDEA) and students with Individual Education Program (IEPs) plans. School divisions were also expected to design professional development to build the capacity of teachers within their division, across divisions, and statewide whenever possible (VDOE, 2014b).

The VDOE gave merit to the idea of incorporating alternative assessments within a *balanced assessment system* at the classroom level by teachers and at the school and division levels by educational leaders to consequently lead to innovation in instruction and deeper learning for students (VDOE, 2014b). The VDOE Assessment Literacy Glossary (2014a) defines a balanced assessment system to recognize how alternative assessments fit within its premise:

> An assessment system that employs multiple types of assessments so that: (1) achievement and growth are taken into account; (2) assessments are matched to learning goals (both core content mastery and skills for success in the modern world); and (3) the need for accountability measures is met, but not at the expense of meaningful information that informs classroom instruction (VDOE, 2014a, p. 1).

On October 27, 2016, the Virginia Board of Education adopted Local Alternative Assessment Guidelines for 2016-2017 through 2018-2019 that replaced the aforementioned 2014-15 Virginia Board of Education adopted guidelines. The updated guidelines clarified the expectation that school divisions were to demonstrate progress in moving toward the use of (performance assessments in their schools (VDOE, 2016b). Subsequently, during the 2016-17 school year, school divisions were expected to use at least one performance assessment in classrooms where a standardized test was replaced with an alternative assessment (VDOE, 2016b). A more recent VDOE (2019b) *Superintendent's Memo, #025-19*, released on January 25,

2019, updated for the second time the Guidelines for Local Alternative Assessments for 2018-2019 through 2019-2020 to replace the 2016-2017 through 2018-2019 version adopted by the Virginia Board of Education in October 2016. The second update follows:

> The updated guidelines clarify the expectation that divisions are to continue the implementation of performance assessments, emphasizing the use of the *Virginia Quality Criteria Review Tool for Performance Assessments* and common rubrics developed by the Virginia Department of Education to achieve consistent expectations for Virginia students. (VDOE, 2018)

The alternative assessment portfolio under examination in this study reflects the most current Guidelines for Local Alternative Assessments.

Local Context

One suburban Virginia school district developed and implemented a *portfolio approach* to assessment in accordance with the VDOE guidelines for alternative assessment. Said district was the site of this study and is called by the pseudonym Performance County Public Schools (PCPS) to protect study participants. PCPS is a large, well-resourced, and demographically diverse school division. PCPS is also diverse in terms of socioeconomic status.

The PCPS portfolio approach most resembles a showcase portfolio type while incorporating features seen in growth and evaluation portfolios. Portfolio types are further explicated in the literature review in chapter two of this study. Teacher-selected performance-based assessments implemented in two middle school social studies courses with a replaced test will be identified as one artifact in a student portfolio and undergo a quality review. A quality review is necessary in order to comply with the most recent VDOE update ensuring alignment of district performance-based assessments to the VDOE Quality Criteria Review Tool for

Performance Assessments. Additional details about the review tool will be provided later in this study.

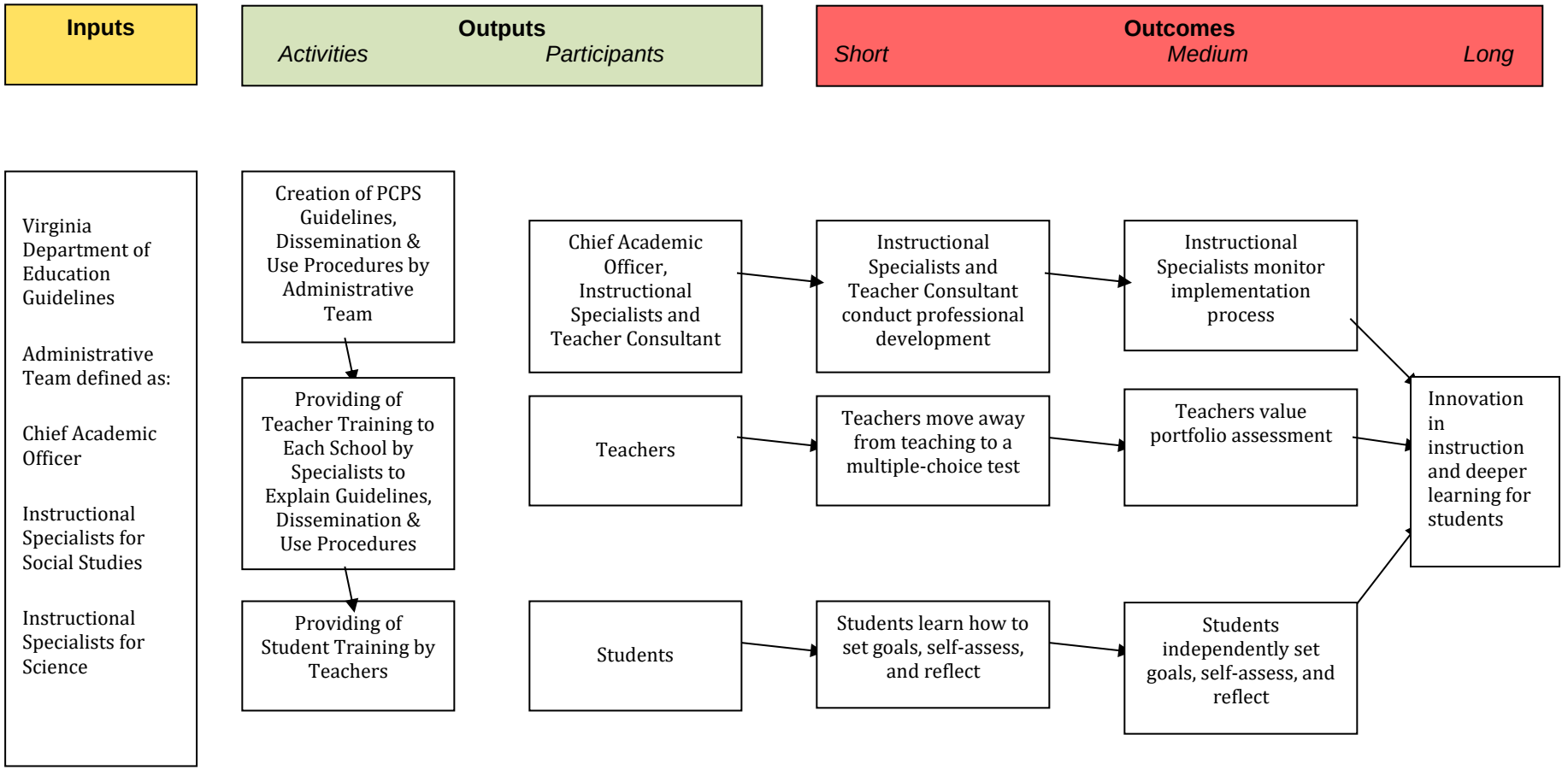
PCPS Guidelines for Local Alternative Assessments for 2014-15 (Appendix B) were developed in response to the aforementioned legislative mandate of House Bill 930 and Senate Bill 306. A combination of several assessment types for each grade and content area were permitted to meet the requirements of the legislation and were designed to provide feedback to parents and teachers regarding the extent to which the student had demonstrated proficiency in the content included in the SOL covered (PCPS). The Chief Academic Office and the Department of Curriculum and Instruction created the PCPS Alternative Assessment Plan. PCPS defined guidelines, to certify local alternative assessments, that recommended to principals and teachers that schools use a *portfolio approach* in order to provide student and teacher choice in the selection of assessment documentation.

Logic Model

Figure 1 (Mertens & Wilson, 2012, p. 56) depicts the logic model showing the sequence of actions that took place in the first years of implementation of the portfolio assessment and projects how the investment in these actions can lead to innovative and deeper learning outcomes for students.

Figure 1

Performance County Public Schools (PCPS) Alternative Portfolio Assessment Implementation Process Logic Model



Inputs

The PCPS alternative portfolio assessment implementation required a variety of inputs. The VDOE guidelines provided guidance to ensure that students were making adequate academic progress in the subject area and that the Standards of Learning was being taught (VDOE, 2014b). The administrative team, consisting of the Chief Academic Officer, Director of Curriculum and Instruction, and Instructional Specialists for Social Studies, Science, and Elementary Language Arts, convened to devise a plan for the alternative assessments that adhered to the VDOE guidelines. The Chief Academic Officer directed the team to develop a portfolio assessment for the eliminated SOL tests.

Outputs

The PCPS administrative team developed the guidelines for the alternative portfolio assessments as follows:

- Writing portfolios are already kept at Grade 5. This serves as authentic evidence of student mastery of writing.
- The portfolios for science and social studies will allow students to set goals for the year and reflect on the goals at the end of the year.
- The portfolios will allow students to choose two or more artifacts that represent the students' best work in each strand.
- Students will self-assess and peer review throughout the year.
- Students will use a Marzano (2003) scale of 1 to 4 to self-assess their work.
- The portfolio will include a menu of recommended formats of evidence that would be acceptable under the state guidelines. Student artifact examples may include, but are not limited to:

- Interactive Notebook
- Reflective Journal
- Content Writing Prompt
- Common Assessment
- Performance-based Assignment
- Project-based Learning Assignment

The following timeline was put into place from August 2014 through June 2015 to implement the PCPS Local Alternative Assessments:

- August 2014
  - PCPS Alternative Assessment Portfolios are printed for each student to use in elementary schools; teachers in middle schools can choose to use electronic portfolios
- September 2014 through June 2015
  - Communication will be shared with parents regarding the purpose and use of the alternative assessment portfolios.
  - Students will choose evidence of their mastery of the content.
  - Students will use a Marzano (2003) scale of 1 to 4 to self-assess their evidence.
  - Students will conduct and participate in a peer review.
  - Teachers will scaffold the selection and self-reflection process for students.
- May through June 2015
  - The Chief Academic Office will require each principal to annually certify that their school has provided instruction and administered an alternative

assessment, consistent with Board guidelines, to students in grades three through eight in each SOL subject area in which an SOL assessment was not administered during the school year.

- Science and social studies instructional specialists will survey teachers for suggestions to improve or enhance the alternative assessment portfolios.

The science and social studies curriculum specialists arranged to visit third grade teams at each of the 38 elementary schools and USI and USII social studies teams at each of the 12 middle schools to provide professional development regarding the implementation of the alternative assessment. A typical professional development meeting included the entire grade level team reviewing a presentation about the implementation process of the alternative portfolio assessment. The presentation communicated the following topics on a variety of slides. Listed below are the topics:

- Purpose for the Local Alternative Assessment
- Expectations from the Virginia superintendent of schools for authentic assessment and instruction
- Mastering learning through traditional and performance-based assessments
- Student goal-setting
- Examples of Artifact of Instruction vs. Artifact of Learning
- Artifacts of student learning exemplars and non-exemplars
- Student reflections
- Performance-based assessments—definition and examples

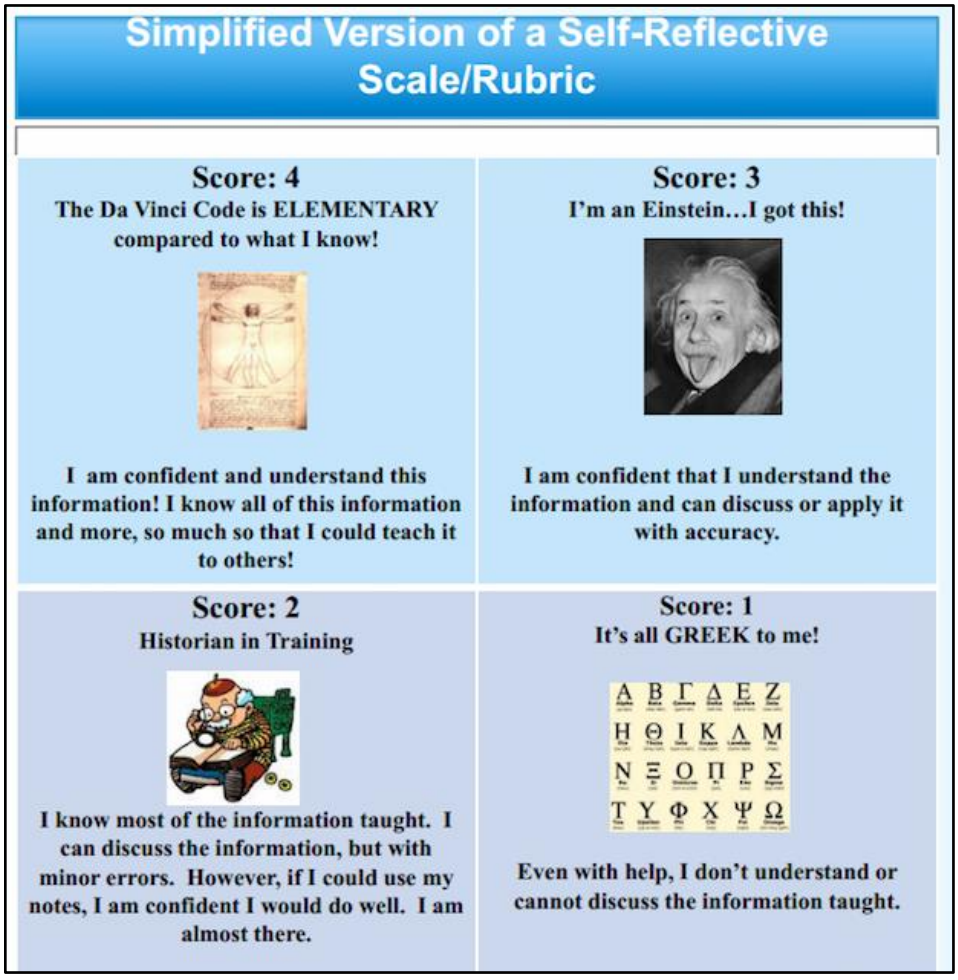
The instructional specialists also shared the same presentation with elementary and middle school principals during a regularly scheduled monthly meeting.

Outcomes

Short Term. Short-term outcomes noted in the alternative portfolio assessment process logic model include students learning how to set goals, self-assess and reflect. Teachers modeled these processes for students. A simplified version of a self-reflective scale was adapted from Marzano (2003) was also introduced, as shown in Figure 2.

Figure 2

Simplified Version of a Self-Reflective Scale/Rubric



Teachers explained to students that they would showcase their best work to classmates by describing why they selected specific artifacts to appear in their portfolio. Teachers were also

encouraged to use performance-based tasks and assessments while simultaneously moving away from teaching to a standardized test.

Medium Term. Intermediate goals include curriculum specialists monitoring the process of the alternative portfolio assessment. Informal surveys of middle school social studies teachers were launched to gauge the progress of the implementation process and to address the needs of teachers and students regarding the process. Within these surveys and informal conversations with teachers and students, what was also monitored was how teachers valued the alternative portfolio assessment and the ability for students to begin to set their own goals. This informal feedback revealed that teachers began to value the opportunity for student reflection based upon their observations of students engaging in the process.

Long Term. A long-term outcome of the alternative portfolio assessment is to see students independently setting their own goals and reflecting upon their work. Another outcome is that students are able to independently self-assess. Finally, in time teachers will value the portfolio assessment and be able to use it as an indicator of student growth. The following section will discuss the potential implications of this study on the future of quality performance assessments available to students.

Significance of Study

The problem statement in the introduction of this study addressed the potential lack of quality performance assessments available to students. Based upon the current shift to a more balanced approach to assessment and instruction in Virginia, there was a need for quality performance assessments. As seen in the research literature, waning teacher capacity in recognizing quality performance assessments could be linked to the absence of professional development in this targeted area due to years of teaching to standardized tests. The readiness in

a sample group of teachers to recognize quality characteristics in performance-based assessments revealed in the findings of this study will potentially guide the development of future professional learning opportunities for teachers. As well, the method and structure of this study will model a process transferrable to other school districts in Virginia and beyond and potentially increase the chances that quality performance assessments are available for all students. The meaning of *quality performance assessments* will be examined later in this study.

Evaluation Questions

To uncover the capacity of teachers to recognize quality characteristics in performance-based assessments and make recommendations to strengthen the quality of performance-based assessments, this study was guided by the following evaluation questions:

1. To what degree do teacher-selected performance assessments used as local alternative assessments to the Virginia SOL tests for accountability meet the state's criteria for quality?
2. What are teachers' perspectives on the merit and the worth of the selected performance assessments?
3. Given the comparative quality of the performance assessments and the perspectives of the teachers, what recommendations can be made to strengthen the use of performance assessments as local alternative assessments to the Virginia SOL accountability tests?

Definition of Terms

This section defines relevant research terms regarding the topic of this proposed study, and consequently appearing in this proposed study.

Alternative assessment - a term used to describe assessments used primarily to determine what students can and cannot do, in contrast to what they do or do not know. In other words, an alternative assessment measures applied proficiency more than it measures knowledge (VDOE, 2014a).

Authentic assessment – performance assessments that focus on embedded curricular ideas that ask students to perform authentic tasks in the context of the subject area (Development of the Assessments, 2014).

Balanced assessment – an assessment that employs multiple types of assessments so that achievement and growth are taken into account (VDOE, 2014a).

Feasibility – the extent to which an evaluation can be implemented successfully in a specific setting (Mertens & Wilson, 2012)

Goal setting – the process of establishing clear and usable targets, or objectives, for learning (Moeller et al., 2012).

Interrater reliability – the degree of agreement among raters who are tasked with scoring a performance task or product (VDOE, 2014a).

Performance assessment or performance-based assessment (PBA) – an alternative method used to measure what a student knows and can do (Kan & Bulut, 2014). It can provide insight to the cognitive process of students and can reveal their attitudes toward the content (Corcoran et al., 2004). A PBA is a means of gathering evidence of student learning that requires students to think at high cognitive levels, undertake a skill-based process, and/or product (rather than select) a response; performance assessments are authentic to the discipline and/or to the real world (Gareis, 2017). A PBA may include a written component, but generally focuses primarily

on the student's demonstration of a specified task and/or the creation of a product and is typically scored using a rubric (VDOE, 2014a).

Portfolio assessment – a means for students to monitor their own learning and communicate their learning and previous experiences to others (Adeyemi, 2015); typically, a systematic collection of student work and artifacts that demonstrate mastery of course and/or content knowledge and skills over an identified period of time (VDOE, 2014a).

Reliability – the consistency of test performance based upon the construction of an assessment and the way it is administered to be free of ambiguous scoring, unclear questions/directions, bias, cheating, or environmental factors (VDOE, 2014a).

Rubric – a description of the criteria for success and levels of achievement for a task or product. Rubrics can be used during instruction to help students maximize and improve the quality of their work and as scoring tools for multiple types of alternative assessments (VDOE, 2014a).

Validity – the degree to which an assessment actually measures the learning it is intended to measure (VDOE, 2014a).

Virginia Quality Criteria Review Tool for Performance Assessments – a set of criteria for the development of performance assessments that measure the application of content knowledge and skills (VDOE, 2018).

CHAPTER 2

REVIEW OF THE LITERATURE

There is reason to believe that performance assessments can enrich learning experiences for all students and provide teachers with useful information to support their understanding of student learning and development (Curry & Smith, 2017). Hence, the purpose of this literature review was to explore empirical, theoretical, and anecdotal research that reported on the varying ways performance assessments were defined, quality characteristics and types, and outcomes they have on teaching and learning. Key topics such as alternative assessment, authentic assessment, performance assessment, performance-based assessment (PBA), portfolio assessment, and performance assessment quality are used to search for literature that directly or indirectly focuses on the topic of this review. Bland and Gareis (2018) noted a lack of available research between the years of 1998 and 2005 regarding the aforementioned key topics, which suggests that teachers today may not be well skilled at recognizing, creating, and scoring quality performance assessments. Thus, the need for building teacher capacity to improve the quality of classroom performance-based assessments was also explored in the review of the literature.

Definitions of Performance Assessment

Performance assessment and *performance-based assessment* were used interchangeably throughout much of the assessment literature to label and define how students demonstrate knowledge and skills. Frey and Schmitt (2010) noted that performance assessments are also sometimes called alternative assessments or authentic assessments. Fundamentally defined, a performance assessment is an alternative method used to measure what a student knows and can

do (Kan & Bulut, 2014). In support of the fundamental definition, a nationally recognized organization known for its advocacy of performance assessments, Project Appleseed (2014), explained that a performance assessment should allow students to demonstrate knowledge and skills but adds the importance of including the process by which students solve problems.

Oberg (2010) stated, "performance-based assessment can be seen as one or more approaches for measuring student progress, skills, and achievement and to think of it as the ultimate form of linking instruction with assessment" (p. 5). Oberg (2010) further noted that the authentic features of PBA can lead students to demonstrate their knowledge, skills, or behavior in a real-life context "rather than the contrived problems [such as selecting from four alternatives to show proficiency] sometimes seen in the classroom setting" (p. 5). Similarly, Koh et al., (2011) stated that performance assessments are intended to measure students' knowledge and skills at deeper levels than traditional assessments and tend to solve authentic or real-world problems. "Real-world problems might include engaging tasks similar to tasks people do in their communities, in businesses, and in scholarly disciplines" (Curry & Smith, 2017, p. 169).

Additionally, Curry and Smith (2017) pointed out that

> performance assessments, however, do not have to be 'authentic' because students can 'perform' tasks that are not meaningful outside of a school context, but for the context of social studies, for example, it is helpful if performance assessments are directly meaningful to their futures as citizens." (p. 169)

Perhaps one of the earliest references to authentic tests was that made by Archbald and Newman (1988) about weaknesses of standardized testing that sought to promote assessment centered on meaningful real-world problems or tasks: "Standardized tests of general academic achievement usually do not provide information useful for improving individual or school performance, and

the forms of achievement they assess usually fall short of most criteria for authenticity (p. 50). Frey et al., (2012) stated that for an assessment to be considered authentic it must have meaning beyond the actual score or grade indicating that the assessment task itself should be meaningful and recognize the 'real-world' of students' values, abilities, and motivations" (p. 12). Frey et al., (2012) further stated that, "This suggests that assessments that require behaviors or cognitive operations that are not intrinsically meaningful, (e.g., responding to multiple-choice questions on externally-produced standardized tests) are not authentic" (p. 13). Additionally, Newman et al., (1996) stated that, "For students, authentic assessment involves *authentic intellectual work* which enables students to engage in higher-order thinking and real world problem solving, rather than just routine use of facts and procedures" (as cited in Koh et al., 2011, p. 139). Gareis (2018) specified that a performance-based assessment is a means of gathering evidence of student learning that requires students to think at high cognitive levels, undertake a skill-based process and/or product (rather than select a response); performance assessments are authentic to the discipline and/or to the real world. Gareis (2017) also pointed out that performance assessments can vary greatly in their uses and forms, as shown in Table 1.

Table 1

Comparing Types of Performance Assessments

Characteristics	Type			
	Constructed-Response	**Stand-Alone**	**Curriculum-Embedded**	**Complex Project**
Number of Intended Learning Outcomes (ILO)	1–2 ILOs	Multiple, subject-specific ILOs	A cogent set of subject-specific ILOs	A complex, integrative set of ILOs & broad aims
Level of Instructional Support from Teacher During Administration	Limited to clarification	Limited clarification & facilitation	Integrated instruction, facilitation, & feedback	Integrated instruction, facilitation, feedback, & guidance
Prescriptiveness of Student Response (i.e., Degree of Student Choice)	Fixed/Convergent (typically little choice)	Convergent (limited choices)	Moderately Divergent (elements of choice in content and/or format of response	Divergent (typically multiple opportunities for student choice)
Appropriate Duration	A portion of a class period (≤ 60 minutes)	1–2 class periods (≥ 60 minutes)	Multiple class periods or days	Multiple weeks or a term

Note. From *PBA Development* [PowerPoint slides], by C. Gareis, March 2017, School of Education, William & Mary.

A portfolio is an authentic performance assessment that can serve as a means for students to monitor their own learning (Adeyemi, 2015). A portfolio that supports learning is in stark contrast to the traditional multiple-choice assessment experience in which selecting one answer can signal the end of learning. Typically, a portfolio is a systematic collection of student work and artifacts that demonstrate mastery of course and/or content knowledge and skills over an identified period of time (VDOE, 2014a). The authenticity of the assessment evolves as learners develop the portfolio by setting learning goals, reflecting on their learning by determining how

they master a set of skills and applied knowledge and attitudes, and explaining reasons why they chose certain works be included in their portfolio (Adeyemi, 2015; Chang et al., 2014). The authentic nature of a learning portfolio is supported by the "means for reflection and critiquing one's own work, evaluating the effectiveness of lessons, and the interpersonal interactions with students or peers" (Adeyemi, 2015, p. 127). Over time, reflections develop into meaningful connections to learning, which is an advantage in helping students retrace and deepen their knowledge and experiences. Furthermore, the act of self-assessment provides students opportunities to take ownership of their learning, because they give input to the assessment process, make choices in determining how they will demonstrate their learning, and make connections between their performance and the learning goal while providing teachers with another source of data for future decisions (Gallavan & Kottler, 2009; Kloser et al., 2017).

Three types of portfolio assessments to consider are growth, showcase, and evaluation. *Growth* portfolios tend to show growth or change over time and reveal students' strengths and weaknesses, thus, emphasizing the process of learning (Mueller, 2018). Students also develop the skills of self-evaluation and goal-setting and are included as a valued and reflective participant in the assessment process. *Showcase* portfolios generally reveal samples of best, favorite, or most important work as determined by the students themselves, thus emphasizing the products of learning. Showcase portfolios are typically prepared and presented to audiences such as teachers, parents, and other students as a culminating exercise to show semester or end-of-year accomplishments (Mueller, 2018). The PCPS alternative portfolio assessment most resembles the showcase portfolio type. Growth and showcase portfolios are frequently implemented within the confines of a classroom and reflect a contextual setting.

Evaluation portfolios are commonly used to document achievement for grading purposes, progress towards standards, and to place students appropriately (Mueller, 2018). Evaluation portfolios may be used on a large-scale performance assessment basis to monitor system performance, program/teacher evaluation, accountability, and broadly defined educational reform (Dunbar et al., 1991). Important to note is that:

> once an assessment portfolio is passed on to next year's teachers, district-wide program assessment, or statewide accountability testing all contextual understanding of limitations that may have existed in the classroom are gone, thus becoming a high-stakes application of educational measurement to the extent that it can affect a wholesale change in a school program affecting all students. (Dunbar et al., 1991, p. 290)

Mueller (2018) described a variety of student work that can be selected as possible samples for respective types of portfolios, as shown in Appendix A.

Portfolio assessments used for the purpose of high-stakes evaluation, as well as other types of portfolio assessments, should undergo a quality review to ensure the most reliable results are communicated. While several types of performance assessments are considered above, added attention was given to the types of portfolios because portfolios are linked specifically to performance assessments at PCPS. Common quality characteristics within a variety of performance assessment types were explored later in this study.

Table 2 shows a side-by-side breakdown by skill and ability; process; and value and attitude of the various ways to describe performance assessments (Bland & Gareis, 2018). *Alternative assessment* was a term used periodically in the literature to denote assessments not associated with the traditional multiple-choice and fill-in- the-blank. This term appears in Table 2 as an umbrella for all others that characterize performance assessment. The common language

throughout suggests the similar nature and purpose of performance, authentic, and portfolio assessments.

Table 2

Summary of Assorted Definitions of Alternative, Performance, Authentic, and Portfolio Assessments Recognized by Various Researchers

	Alternative Assessment Type		
Criteria	**Performance**	**Authentic**	**Portfolio**
SKILL & ABILITY	Measures skill or ability (Frey & Schmitt, 2010) and clearly articulates what teachers should be teaching and what students should be learning and allows for evaluation of both process and product (Lai, 2011).	Demonstrates knowledge, skills, or behavior in a real-life context rather than contrived problems for the classroom setting (Oberg, 2010) and performance assessments that focus on embedded curricular ideas that ask students to perform authentic tasks in the context of the subject area ("Development of the Assessments", 2014).	Shows evidence of mastery of a set of skills, applied knowledge, and attitudes (Adeyemi, 2015; Chang et al., 2014).
PROCESS	Demonstrates knowledge and skills, including the process by which they solve problems (Project Appleseed The National Campaign for Public School Improvement, 2014) and provides insights to students' cognitive process and reveals attitudes toward content (Corcoran et al., 2004)	Poses an intellectually interesting and personally meaningful problem or task and engaged higher-order thinking (Frey et al., 2012; Koh et al., 2011) and are authentic to the discipline and/or the real world (Bland & Gareis, 2018)	Evaluating the effectiveness of lessons of interpersonal interactions with students or peers (Adeyemi, 2015) and scaffolding portfolio types over time (Duffy et al., 1999)
VALUE & ATTITUDE	Has value, interest and is motivating to students beyond the classroom into the "real-world" (Frey et al., 2012; Lai, 2011) and can be described as the tasks people do in their communities, in businesses, and in scholarly disciplines (Curry & Smith, 2017)	Realistic because the questions, tasks, or problems have value and interest beyond the classroom into the "real-world" of students' values, abilities, and motivations (Frey et al., 2012; Koh et al., 2011) and can be described as the tasks people do in their communities, in businesses, and in scholarly disciplines (Curry & Smith, 2017)	Artifacts students select to go into the portfolio provide students a real-world opportunity (Adeyemi, 2015) and the act of self-assessment provides students opportunities to take ownership of their learning (Gallavan & Kottler, 2009; Kloser et al., 2017)

Note. From "Performance assessments: A review of definitions, quality, characteristics, and outcomes associated with their use in K-12 schools," by L. Bland & C. Gareis, 2018, *Educators' Journal, 11*, p. 7.

The following section suggests ways the defined performance assessments could function by providing authentic examples of common characteristics found in performance assessments. Examples include student motivations, as well as alignment to skills and content.

Common Quality Characteristics within a Variety of Performance Assessments

Quality performance assessments have common characteristics that include:

- asking students to perform, create, or produce something;
- tapping higher-level thinking and problem-solving skills;
- using tasks that represent meaningful instructional activities;
- involve real-world applications; and
- using human judgment to do the scoring. (Corcoran et al., 2004, p. 1)

VanTassel-Baska (2013) asserted that performance assessments themselves serve as a basis for continued, authentic learning for students, including gifted students. Writing assignments, such as essays, are noted by some teachers as one of the most popular types of performance assessments used in classrooms (Frey & Schmitt, 2010). Essay type assessments can be considered a performance assessment when their purpose is to measure skill or ability and be subjectively scored (Frey & Schmitt, 2010). Bland and Gareis (2018) contended that subject-specific knowledge on the part of the teacher is necessary when subjective scoring is employed. Corcoran et al., (2004) suggested journal entries as a meaningful performance assessment because "if it is tied to teaching and learning objectives, the teacher can gain insight into students' cognitive progress and reveal their attitudes toward content" (p. 214). Thus, performance assessments can be used to monitor student growth (Bland & Gareis, 2018).

Variations of performance assessment designs that enable students to demonstrate knowledge and skills can include learning logs, posters, experiments, debates, mock interviews,

artistic work, writing/performing music, and/or dance (Oberg, 2010). Another option includes historical writing assessments that evaluate the skills needed to support historical reasoning, which include analytical and evaluative thinking (Bland & Gareis, 2018). The analytical thinking skills involved in historical writing assessments include being able to examine evidence; weigh conflicting accounts; consider biases; construct arguments grounded in evidence; and prepare students to understand the complexities of our social world, evaluate information responsibly, and ask difficult questions (Monte-Sano, 2008). Educators who use historical writing as a performance assessment give students another avenue to develop their literacy skills and apply them in a way authentic to the discipline (Bland & Gareis, 2018). The document-based question (DBQ) is noted as one type of historical performance-based writing assessment. DBQs engage students to use thinking and writing skills to analyze text, evaluate the credibility of sources, read for understanding, and use evidence from the text to justify their response to a lesson designed as a module or unit (Johnson, 2016).

Curry and Smith (2017) referenced several examples of performance assessments that support three important aspects: "demonstration of knowledge, ability to reason, and ability to communicate conclusions" (p. 169). Such assessments are published by the National Council for the Social Studies and include these examples:

- *The Human Impact of Natural Disaster*s, wherein students act as members of the United States Government responsible for approving new chemical manufacturing plants, tasked with ensuring the accidental spills and leaks do not happen (Pang, 2010).

- *A Digital Age* inquiry, in which students work together to create a civilization from the ground up, using technology for research, collaboration and dissemination (Bennett & Berson, 2007).

Notable work with performance assessments can be seen at the national level in Singapore with Thinking Schools. In 1997 the Singapore Ministry of Education launched the vision *Thinking Schools Learning Nation* to develop creative and critical thinking skills, a lifelong passion for learning, and nationalistic commitment in the youth of the country (Koh et al., 2011). "Teachers are encouraged to expand their repertoire of teaching and learning strategies to include new and innovative pedagogies, communicate effectively, collaborate widely, and solve problems reflectively" (Koh et al., 2011, p. 136). Thinking Schools Learning Nation involves teachers moving away from more traditional approaches to teaching and learning, such as rote memorization, to authentic assessment. Koh et al., (2011) point out that changes in classroom assessment practices are necessary to promote thinking schools. The mode of day-to-day classroom assessment practices must foster engaged learning and enhance students' mastery of 21st-century competencies. Done well, these assessments reflect more intellectually challenging learning goals and include more authentic, open-ended assessment tasks such as "sustained written prose where students are asked to elaborate on their understanding, explanations, arguments, and/or conclusions" (p. 140).

Bland and Gareis (2018) recognized the formative purposes of performance assessments when citing Oberg (2010), who advocated for the use of performance assessments as a pre-assessment alternative to the traditional quick and simple paper-pencil methods to support teachers in their instructional planning. Pre-assessments administered in an authentic format provide teachers an important glimpse into the prior knowledge of students' skills and interests

to help develop high quality and effective curriculum and instruction (Oberg, 2010). Oberg (2010) suggested teachers consider the following guidelines when developing performance-based pre-assessments:

1. What to understand about the lesson and student learning that the teacher wants to know if the student knows;
2. How students can demonstrate current knowledge in a unique or non-standard way;
3. What the criteria are for competence and mastery of the content;
4. How to judge student competence;
5. How to provide feedback in a constructive manner;
6. How to include the student within this process; and
7. How the results will be used to guide instruction and differentiation (p. 6).

These guidelines can support teachers when creating performance pre-assessments. Consider a "mock" store in the classroom to observe how well students can count money when purchasing items as well as when receiving or giving change (Oberg, 2010). "Teachers observe adding, subtraction, multiplication skills as well as problem solving skills, language skills, and social interactions" (Oberg, 2010, p. 6). Using this performance and a teacher's observation checklist, a pre-assessment of authentically applied understandings and skills related to the intended learning of fundamental financial literacy can be obtained (Bland & Gareis, 2018). Oberg (2010) added that students with special needs or students learning English as a second language may benefit from performance-based pre-assessments as an alternative avenue for demonstrating their skills and knowledge, thus providing a pathway to success in the classroom. Traditional assessments tend generally to serve as a prohibitive tool to show what students with specialized needs know and can do (Oberg, 2010).

Adeyemi's (2015) study of middle school students demonstrated that portfolio assessments authentically contribute to the teaching and learning process. Each of these assessment types involves students in the process, thus giving them a more meaningful role in improving achievement:

> Learners' achievement and their feelings of responsibility for monitoring their own progress provide an intrinsic motivation of interest in a task. Not only do the students in this study relate the different aspects of the information to one another, they also connect them to their previous learning and personal experiences. (Adeyemi, 2015, p. 131)

Additionally, utilizing a variety of assessment tools (as made available through portfolio and other types of authentic assessments) to gain insight to what students actually understand can inform teachers' instructional decisions and the feedback they provide to their students (Kloser, et al., 2017).

Barber et al., (2015) documented a measure of performance assessment known as "Digital Moments." This performance assessment was conducted in an undergraduate course in a university; however, the concept can also apply to high school courses that use online synchronous and asynchronous methods of instruction. In this example, university students enrolled in the course "Psychological Foundations and Digital Technology" submitted one reflection per week, identified as a "Digital Moment," to an online learning management system (e.g., Blackboard) based on three hours of podcasts. Students were encouraged to submit original, creative, and inventive reflections. Exemplars included words, phrases, pictures, colors, and musical links (Barber et al., 2015). Students viewed their classmates' work and gained new ideas regarding ways to complete their assignments and worked collaboratively to develop tools to evaluate their work and the work of others. "Students were taught how to give and receive the

type of feedback that stretched beyond ‘great job’ that is essential for moving the project forward” (Barber et al., 2015, p. 63). “Digital Moments” is a strategy to create a professional learning community and a format for students to use problem based learning strategies and to authentically assess their learning (Barber et al., 2015):

> It allows for many of the parameters in authentic learning environments to exist. Students learn in authentic contexts, do tasks of their choosing, collaborate with others, and have access to peers who share expertise in the particular technology they wish to learn. This creates collaborative construction of knowledge, coaching and scaffolding, and embeds assessment within the learning process. (p. 63)

In general, high-quality performance assessments are more motivating to students, clearly articulate what teachers should be teaching and what students should be learning and allow for evaluation of both process and product (Lai, 2011). The next section described various effects of performance assessments on students. The examples will reflect outcomes connected to deeper learning.

Outcomes Associated with Performance Assessments

The outcomes associated with the use of performance assessments generally reveal that the depth of knowledge and types of skills developed by students are greatly influenced by the nature and format of the assessments (Darling-Hammond & Adamson, 2010). An outcome reached in the previously mentioned undergraduate course in a study of digital pedagogy saw students exhibit greater competence and confidence in using digital open resources (Bland & Gareis, 2018). Students felt empowered to take the reins of their own learning and needed less direction from their instructor. Additionally, learners developed autonomy, engagement, and motivation; self and peer assessments grew to be more meaningful; and there was a shared

development of collective knowledge (Barber et al., 2015). Participating in courses that embrace authentic, performance-based opportunities to learn helps build a foundation from which authentic assessment, student ownership of learning, and peer support can occur in an ongoing way as learners make the important shifts in power to owning their learning and becoming problem-based inquirers (Barber et. al., 2015). For instance, when learning social studies, authentic and portfolio assessments can "go a long way to ensuring academic success in learners as they are vital tools in testing and teaching for both teachers and students" (Adeyemi, 2015, p. 131).

Comer (2011) asserted in her article about young adult literature and alternative assessment measures that performance or authentic assessments engage students metacognitively. Additionally, Comer (2011) stated that there is more opportunity for students to pull from knowledge they already have to answer questions. Corcoran et al., (2004) wrote about students being more engaged and more willing to assess their own learning, and state that creative student projects reflect a range of intelligences. These things occur when students perform, create, produce, or do something involving higher-level thinking and real-world applications. Hallam et al., (2007) studied the effects of outcomes-driven authentic assessment on classroom quality. Their study suggests that an authentic assessment approach, which includes a performance-based curriculum-embedded assessment approach, may have a positive impact on the language and literacy environment.

Another study with a positive impact on the literacy environment was Johnson's (2016) study of the DBQ Project. This study provided evidence to show that the DBQ Project instructional strategies utilized in the implementation of document-based questions can improve student writing of both short and extended constructed responses. Positive outcomes are linked to

the use of DBQs in Aiken County, South Carolina and Capistrano Unified School District in California. In 2017-18, the first year of DBQ implementation in Aiken County, seventh-grade state assessment scores in social studies increased from 57.1% passing to 66.7% passing (Johnson, 2016). According to the district leadership, this 9.6% increase was credited to the instructional shift in the use of the DBQ by using a six-step approach developed by The DBQ Project Method™. Further, in a controlled study with the entire district Grades 6-12, Capistrano Unified School District in 2017-18 saw significant growth in the number of proficient essays (score = 3), as well as high-proficient essays (score = 4) by implementing the six-step approach (Johnson, 2016).

Table 3 summarizes previously mentioned select examples of performance assessments and outcomes associated with their use (Bland & Gareis, 2018). The examples were selected because they expressly incorporate a performance task.

Table 3

Summary Samples of Performance-Based Assessment (PBA) Types, Characteristics, and Outcomes

PBA Type	Researchers	Characteristics of PBA	Observed Student Outcomes
Essay	Frey & Schmitt (2010)	Essay type assessments can be considered a performance assessment when their purpose is to measure skill or ability and be subjectively scored	Increased learning in the classroom as well as increased test scores
Journal	Corcoran et al., (2004)	Writing tied to the learning objectives and use human judgment to complete the scoring	Insight into cognitive process of students and attitudes toward the content
Historical Writing	Monte-Sano (2008)	Writing tied to analysis of evidence, weighing conflicting accounts, determining bias, constructing arguments, and asking difficult questions	Development of literacy skills and application in an authentic way
Document-based Question (DBQ)	Johnson (2016)	Students evaluate primary and secondary sources, analyze and evaluate their importance, and take a position and defend a point of view of their own.	Improvement of both short constructed responses and extended constructed responses to show understanding
'Thinking Schools'	Koh et al., (2011)	Communication, collaboration, and problem-solving pedagogies that support authentic assessment such as written prose where students are asked to elaborate on their understanding, arguments, and/or conclusions	Development of creative and critical thinking skills and a lifelong passion for learning and nationalistic commitment in the young
Pre-assessment	Oberg (2010)	A mock store in the classroom to observe how well students can count money when purchasing items as well as when receiving or giving change	Teachers observe adding, subtraction and multiplication skills as well as problem solving skills, language skills, and social interactions in an authentic format to help develop high quality and effective instruction and curriculum; especially effective for special needs students and English Language Learners
Portfolio	Adeyemi (2015)	Students are involved in the process of monitoring their own learning and communicating their learning and previous experiences to others.	Intrinsic student motivation of interest in a task and feelings of responsibility for monitoring their own progress which gives students a meaningful role in improving achievement
Digital Moments	Barber et al., (2015)	Digitally, students complete tasks of their choosing, give and receive feedback to self and others, and immediately apply knowledge to move their project forward.	Effective use of feedback develops valuable online collaboration and communication skills and embeds assessment within the learning process.

As evidenced in the Table 3 summary, there is a wide range of performance assessment types. Outcomes associated with their use in Table 3 appeared to reflect growth in the application

of skills-based learning, an intrinsic motivation to learn, and a heightened awareness of the manner in which teachers and students valued performance assessments. The impact of rubrics and the vital role they play in the implementation of performance assessments must not be overlooked; rubrics are the tools used to distinguish successful student outcomes. The next section stressed the importance of sustained professional development in designing and implementing quality authentic assessments and rubrics.

Professional Development

Bland and Gareis (2018) earlier noted the scarcity of performance-based assessment research studies between the years of 1998 and 2005. This lack of research may suggest that today's teachers are not well prepared to create, implement, and score performance-based assessments. A case in point reveals results from a 3-year longitudinal study on assessment practices in social studies classrooms and notes that study participants, as recent as the years 2013-2015, "were often more likely to report the use of assessments of knowledge (including selected-response items) than performance-based assessment techniques" (Curry & Smith, 2017, p. 168). Curry and Smith (2017) "reinforced the need for professional development that helps teachers see how performance-based assessments can be used to boost student performance on high-stakes assessments." (p. 168). Teachers often rely on assessments created mainly by others, such as textbook companies, and therefore may benefit from professional development and support on best practices related to performance assessments (Frey & Schmitt, 2010). Monte-Sano (2008) maintained that many teachers do not have the prerequisite skills necessary to teach performance-based tasks, such as evidence-based historical thinking and writing, and that professional development is needed.

The need for professional development to support evidence-based historical writing was noted by teachers in the DBQ Project study as a key to successful DBQ implementation. One teacher said, “I don’t think my first attempt at completing a DBQ module in the classroom would have been as successful if I had not gone through the module process myself first. That was the most helpful aspect of the training to me” (Johnson, 2016, p. 42). Koh et al., (2011) described a move toward performance assessment at a national level in Singapore, stating that in order for this shift to occur, professional development for teachers in the practice and use of authentic assessment is essential. The findings in this two-year empirical study reveal that:

> Ongoing and sustained professional development in designing and implementing authentic assessments and rubrics was more effective than ad-hoc, one- or two-day workshops to build teachers’ capacity in improving the quality of classroom assessment tasks in English, science, and mathematics. As a result, there was also significant improvement in the quality of student work in response to the high intellectual demands of the assessment tasks. (p. 144)

Coupled with the implementation of performance assessments were the use of rubrics to reflect on and evaluate student productivity, performance, and products. Rubrics are essentially scoring guides that formalize the evaluation process to provide clear and fair results to students; rubrics delineate the teacher’s expectations for performance (Oberg, 2010). A study conducted by Kan and Bulut (2014) examined the effects of teacher experience and rubric use in performance assessments. In this study, eighth-grade students were given a performance task, and 17 teachers with and without a rubric graded their responses. The performance assessment graded first without the use of a rubric resulted in an inconsistency of scores among the teachers. When the assessments were graded several weeks later using a rubric, the consistency among the

scores given by the teachers increased, thus supporting the use of rubrics when grading performance assessments. Clearly developed and aligned rubrics, established to assess the knowledge and skills of students, are essential when considering the overall process of quality performance assessment implementation.

Implementing quality performance assessments requires knowledge of their design. Professional development for teachers to help recognize the quality components of performance assessments is key, as well as best practices in the use of rubrics to score student work. According to the aforementioned research studies, there has been an opportunity time gap for teachers to engage in professional learning regarding best practices in the implementation of performance-based assessments. Therefore, in an effort to support the needs of teachers to best support students, one aim of the current study is to develop targeted professional learning opportunities based upon the readiness of PCPS teachers to recognize and implement quality performance-based assessments. The following section will recap the highlights presented in the literature review and provide suggestions for ensuring that people graduating from universities with education degrees are prepared to implement quality PBAs.

Conclusion

The purpose of this literature review was to explore empirical, theoretical, and anecdotal research regarding the various ways performance assessments were defined, quality characteristics and types, outcomes related to their use, and evidence supporting the need for teacher professional development. This review noted that performance assessments could be defined in several ways. A fundamental definition was earlier identified as: *an alternative method to measure what a student knows and can do* (Kan & Bulut, 2014). However, due to the numerous and varied examples of ways performance assessments are viewed and utilized in the

literature, the fundamental definition does not truly account for the robust learning experiences that can occur when performance assessments are implemented. Table 4 shows the frequency of common words and concepts associated with performance assessments by various researchers that appeared in Tables 2 and 3 (Bland & Gareis, 2018, p. 18).

Table 4

Content Analysis Summary of Common Words and Concepts Associated with Performance Assessments by Researchers from Tables 3 and 4

Words and Concepts	***f***
Skills/Abilities	12
Authentic or Real-world Contexts	11
Tasks	9
Value/Attitudes	9
Knowledge/Understanding	7
Process	6
Social/Emotional	5
Subjective Scoring	2

By considering the words and phrases in Table 4 and the frequency of occasions they appear in the existing literature, a richer description would be that, "performance assessments, considered as a task or product, reveal the ability of students to authentically demonstrate knowledge, skills, and processes in a way that provides value, interest, and motivation to students beyond the actual score or grade" (Bland & Gareis, 2018, p. 18). Bland and Gareis (2018) supported the view that additional research in the area of performance assessment will be beneficial as the United States moves from the standardized testing era of NCLB (2001) to ESSA (2015). Bland and Gareis (2018) also agreed that not all standardized tests will likely be abolished under ESSA, but there will be a move away from a "one size fits all" assessment and more flexibility given to Local Education Agencies (LEA) to assess students in a variety of ways (NCLB, 2001).

References to building the capacity of teachers in the use of performance assessments with sustained professional development are made throughout the literature. Additionally, institutions preparing teachers for the field of education should consider revising coursework and student teaching experiences to prepare beginning teachers for this shift in assessment practices (Koh et al., 2011).

CHAPTER 3

METHODS

This chapter presented the purpose and design of the study. The purpose of this study was to find characteristic features of quality in performance-based assessments (PBA) in a sample population. Key components of the study design were research questions, participants involved, data sources and collection methods, and data analysis methods. Ethical considerations and assumptions, delimitations, and limitations were also included as elements of the design. Additional detailed information relevant to the study was referenced and available in appendices. Findings were applied to improving the quality of PBAs in Performance County Public Schools (PCPS).

Evaluation Questions

For the purpose of this study, teacher-selected performance assessments were obtained and examined from PCPS social studies courses *United States History to 1865* and *United States History 1865 to the Present*, respectively. The courses are often referred to as "USI" and "USII." As stated in Chapter 1, the following evaluation questions will guide this study:

1. To what degree do performance assessments used in PCPS as local alternative assessments to Virginia Standards of Learning (SOL) tests for accountability meet the state's criteria for quality?
2. How are performance-based assessments being implemented in USI and USII classes in PCPS?

3. What are USI and USII teachers' perspectives on the merit and worth of the performance assessment process in PCPS?
4. What are USI and USII teachers' recommendations for developing and using quality performance assessments in PCPS moving forward?

The evaluation questions focused on understanding the problems, strengths, and weaknesses of the implementation process and the way affected people valued performance assessments as an alternative assessment and they can lead to improvement of the process (Stufflebeam & Shinkfield, 2007).

Design

A responsive program evaluation approach guided the implementation of this mixed-method study in search of performance assessment quality. A key tenet in Stake's model of responsive program evaluation was receptiveness to the concerns of stakeholders. Therefore, information collected in this study was "open to acknowledging common and single, one-of-a-kind insights that could potentially change the perception of quality" (Stake, 2004, p. 88) in a performance assessment. "While an outcome of responsive program evaluation may be to eventually alleviate or remediate, or develop or aspire, the purpose is mainly to understand" (Stake, 2004, p. 89). What is valued in the approach is being able to weigh the evidence, make judgments, and report the different ways in which quality and limitations are understood. The approach used both *criterial measurement* and *interpretation* to respond to the issue of quality and "specifically draw attention to what educators are doing and less attention to what students are doing" (Stake, 2004, p. 101).

The criterial measurement tool used in this study was the *Virginia Quality Criteria Review Tool for Performance Assessments* (QCRT; Appendix C). The QCRT is a "set of criteria

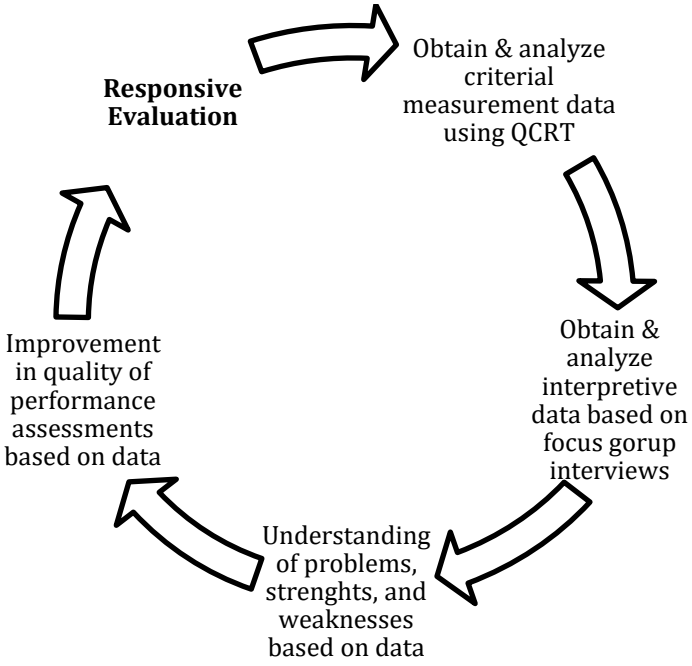
for the development of performance assessments that measure the application of content knowledge and skills and is designed to support comparability in rigor and quality across the state" (VDOE, 2018). Holli Cook, coordinator of assessment at the VDOE, noted that, "What was unique was the Virginia General Assembly making the development of the alternative assessment a local requirement" (H. Cook, personal communication, September 3, 2019). Cook also shared that the VDOE had to walk a fine line by developing a flexible enough tool for local divisions to fit the local context but still meet VDOE guidelines for locally developed performance assessments. With a generous grant from the Hewlett Packard Foundation, members of the VDOE, along with Dr. Christopher Gareis, Professor of Educational Leadership at the College of William and Mary, and Diane Washenburger, Director of Instruction for Salem City Schools attended a conference at the Stanford Center for Assessment, Learning and Equity (H. Cook, personal communication, September 3, 2019). The purpose of the Stanford Center for Assessment, Learning and Equity is to provide "technical consulting and support to schools and districts that have committed to adopting performance-based assessment as part of a multiple-measures system for evaluating student learning and measuring school performance" (n.d.). The conference was considered a jumping off point in the development of the QCRT. Later, input from educators representing each of the eight Superintendent's Regions was instrumental in the development of the Virginia tool during collaborative meetings facilitated by Dr. Gareis. (H. Cook, personal communication, September 3, 2019). Sponsored by the Virginia Association of School Superintendents, the first explanation and iteration of a Virginia-specific quality criteria tool took place in January 2017 at a statewide workshop for teachers and educational leaders.

The criterion-related evidence based on the QCRT revealed characteristics of quality in PBAs currently implemented with students in PCPS. An example of the implementation process

of the QCRT is presented in the data collection section later in this chapter. The interpretive data used in this study included feedback collected from focus group discussions and included participants' personal observations about the implementation process and value of the performance assessments. With the consent of focus group members, conversations were recorded, transcribed, coded, and analyzed for revelatory insights. Results gleaned from this study will be used to advance the quality of performance assessments in PCPS. The flow chart in Figure 3 summarizes the Responsive program evaluation approach used in this study.

Figure 3

Responsive Program Evaluation Approach Summary Flow Chart



Note. Quality Criterion Review Tool (QCRT)

Participants

A total of 29 teachers participated in focus groups from 10 of the 12 middle schools. As

an alternative to the accountability SOL tests, said teachers currently implement performance assessments as mandated by *Virginia Code §22.1-253.13:3,* described in Chapter 1.

Role of Evaluator

As the evaluator in this study, I acted as a facilitator of change as formative evaluations were made to improve the quality of performance-based assessments in PCPS (Mertens & Wilson, 2012). I was also the secondary social studies curriculum specialist for PCPS. Although I was not responsible for supervision or evaluation of these teachers, I had established relationships with them that could be characterized as being supportive, fair, caring, trusting, reliable, and professional. Based upon my established relationship with participants of five or more years, I was able to provide relevant insights into the context of the study and use helpful patterns of communication with participants throughout the life of the study (Mertens & Wilson, 2012).

I was aware of the potential pressure study participants may have felt to volunteer for this study and made every effort to help them feel at ease to speak with candor. I placed great value on open and honest responses from participants and protected the integrity of each teacher to respond without fear of retribution. This issue is also addressed in the section entitled "Ethical Considerations."

Quality Review Team

A quality review team examined USI and USII performance-based assessments for features of quality. The team consisted of the secondary social studies curriculum specialist and the PCPS elementary social studies curriculum specialist. Both of us were professionally licensed in the state of Virginia and certified in secondary social studies content knowledge and academic skills by The Praxis® Test. As well, we attended and received certificates of

participation from VDOE sponsored workshops where training occurred in the implementation of the QCRT. The procedures implemented during said workshops were used to review PCPS USI and USII performance-based assessments for quality, as described below.

The VDOE workshops used a collaborative workshop approach facilitated by the instruction and assessment departments at the VDOE by engaging small groups of participants to examine and discuss each of the seven criteria listed in the QCRT. Each criterion appeared in a row with a column designated for a rating of zero to three and a column designated for written evidence to substantiate the rating. Table 5 is a sample row and quality rating scale rubric as seen in the actual QCRT. The rubric for the quality rating is as follows: 0 = *No Evidence*; 1 = *Limited Evidence*; 2 = *Partial Evidence*; 3 = *Full Evidence*.

Table 5

Virginia Quality Criterion Review Tool (QCRT) Sample Row 1 for Criterion 1

No.	Description	Quality Rating	Evidence or Rationale
1A	Virginia Standards of Learning (SOLs) selected for the performance assessment are clearly listed in a task template, developmentally appropriate for target students, and aligned to the grade-level scope and sequence or grade-level curriculum. Performance assessment components, resources/materials, and student products are aligned to the listed SOLs.		

During the VDOE workshop, participants were provided with a sample performance assessment to compare to each criterion listed in the QCRT. During the comparison process, participants were first instructed to individually rate and justify the rating of the performance assessment and then take turns sharing ratings and written justifications with group members. Raters with scores not in agreement discussed their rating justifications until there was some level of consensus (within at least 1 point). This process built agreement between raters for the purpose of distinguishing quality aspects of performance assessments and further explored in the

data analysis section of this study.

Data Sources and Collection

Two data sources helped answer the research questions. The first data source was PBAs administered to USI and USII middle school social studies students and reviewed for quality using the QCRT. The second data source was school-based focus group data regarding the value, use, and process of said performance assessments and performance assessments in general.

PBAs

The primary data source in this study was sample teacher-selected performance-based assessments implemented by approximately 58 USI and USII social studies teachers from the 12 middle schools in PCPS as part of an alternative approach to assessment. Each performance-based assessment was noted as one artifact in a collection of assessments within a portfolio maintained by students and aligned to the skills and content required by the state curriculum in USI and USII (Appendices D and E). Some middle school grade-level social studies teachers implemented common performance assessments whereby all teachers on a grade-level team implemented the exact same performance assessment with their students. As a result, at least one performance assessment from USI and at least one performance assessment from USII was reviewed for quality from each of the middle schools with common assessments. One performance assessment from each USI teacher and each USII teacher was reviewed for quality from schools that did not implement common performance assessments. I reviewed 24 performance assessments for quality assurance.

Focus Groups

The second data source in this study was focus group meetings composed of teachers from across the 12 middle schools in PCPS. I requested that the social studies department leads

at each of the 12 middle schools share an email (Appendix F) with their department members; this asked USI and USII teachers to volunteer via a Google form to participate in a focus group meeting lasting approximately 75 minutes. I scheduled multiple focus group meetings after teachers' contractual workday at a time and location convenient for the teachers. Questions asked during these discussions supported the research questions concerning teachers' knowledge and experiences with performance assessments. With the permission of each participant, discussions were digitally recorded and transcribed. Participants were made aware that recordings were utilized solely by me for the purposes of this study and disposed of at the conclusion of the study. Focus group discussions commenced with the introduction and set of questions available in Appendix G. The two data sources used in this study to answer the four major research questions are summarized in Table 6.

Table 6

Table of Specifications for Data Sources

Research Question	Data Source	Data Analysis
1. To what degree do performance assessments used as local alternative assessments to Virginia Standards of Learning accountability assessments meet the state's criteria for quality performance assessments?	Teacher-selected performance assessments	Descriptive statistics (e.g., *Mean, Median, Mode Standard Deviation, Range)* Content analysis of "evidence" comments written by reviewers to justify ratings, coded by strengths and weaknesses using the Virginia Quality Criteria Tool for Performance Assessments Content analysis of answers to open-ended questions
2. How are performance-based assessments being implemented in USI and USII classes in PCPS?	Focus groups	Content analysis of answers to open-ended questions
3. What are USI and USII teachers' perspectives on the merit and worth of the performance assessment process in PCPS?	Focus groups	Synthesis of data sources
4. What are USI and USII teachers' recommendations for developing and using quality performance assessments in PCPS moving forward?	Integration of findings from focus groups	

Note. United States History to 1865 (USI); United States History 1865 to the present (USII); Performance County Public Schools (PCPS)

Data Analysis

This study used a pragmatic approach, which "focuses primarily on data that are found to be useful by stakeholders and advocates for the use of mixed methods" (Mertens & Wilson, 2012, p. 88). The constant comparison data analysis strategy allowed me to continually interact with and code relevant emerging data pertinent to the topic under study. Marginal notes became the basis for creating categorical codes for use in the analysis of the qualitative data (Stufflebeam

& Shinkfield, 2007). "*Open* and *axial codes* were compared and contrasted in order to refine and make meaning of the research data" (Mertens & Wilson, 2012, p. 453). Because "codes are the building blocks of qualitative analysis, a codebook was developed that included each potential code and a brief description of the meaning of that code" (Mertens & Wilson, 2012, p. 445). Examples of codes and brief descriptions were provided later in this study.

PBAs

The QCRT is a rating scale designed to measure the quality of performance assessments (see Appendix B). This tool contained seven criteria and matched against the context and design of a performance assessment. Each criterion was rated on a scale of 0–3, based on how well it was represented in the assessment as follows:

- 0 – No Evidence
- 1 – Limited Evidence
- 2 – Partial Evidence
- 3 – Full Evidence

The QCRT provided a designated frame where written evidence was included for the purpose of substantiating or justifying the rating for each criterion. Because two raters reviewed each performance assessment for quality, a statistical procedure known as *percentage of agreement* was used to establish interrater reliability. To establish interrater reliability in this study, the two quality review team participants independently rated the quality performance assessments using the QCRT. Then, the quality review team engaged in a practice known as *comparative consensus* by stating their individual ratings for each criterion and noting the cited evidence necessary to support the ratings. Ratings that did not agree were deliberated until there is a consensus rating of 100% agreement. The process to determine percent agreement for two raters was as follows:

1. Count the number of ratings in agreement (e.g., 13).
2. Count the total number of ratings (e.g., 17).
3. Divide the total by the number in agreement to get a fraction (e.g., 13/17).
4. Convert to a percentage (e.g., 13/17 = 76%).

Sample Row 2, shown in Table 7, demonstrated a hypothetical quality rating and evidence justification of Criterion 1: Standards/Intended Learning Outcomes. Based upon the quality rating and rationale in the provided example, the performance assessment was revised to align to the targeted SOL. Each of the seven criteria in the QCRT was used with the same quality check procedure as seen in Table 7.

Table 7

Virginia Quality Criterion Review Tool Sample Row 2 for Criterion 1

No.	Description	Quality Rating	Evidence or Rationale
1A	Virginia Standards of Learning (SOLs) selected for the performance assessment are clearly listed in a task template, developmentally appropriate for target students, and aligned to the grade-level scope and sequence or grade-level curriculum. Performance assessment components, resources/materials, and student products are aligned to the listed SOLs.	0	There is no evidence in the PBA to show alignment to a targeted SOL and therefore lacks a developmentally appropriate target for students.

Note: PBA = Performance- Based Assessment

I carefully read each QCRT criterion rationale and used an *emergent thematic coding process* by implementing the following steps:

1. School data was organized alphabetically and assigned number codes 1-12 to protect the identity of the school.
2. Each school's QCRT was labeled with distinct codes.
3. Each of the seven criteria, along with related subsections, from each QCRT was

copied and pasted into unique documents and identified by respective codes.

4. I carefully read all information and made marginal notes to highlight specific words and phrases that were relevant to the purpose of the study.
5. Summaries were created from the marginal notes for each criterion and subsequent subsections to help develop the final report.

Focus Groups

Audio-recorded discussion sets from focus groups were transcribed by a professional transcription service and analyzed in the following steps:

1. School data were organized alphabetically and labeled with numbers 1-12 to protect the identity of each school.
2. Each school transcription set was labeled with distinct codes by focus group.
3. I used an *emergent thematic coding process* during careful reading analysis of each discussion set transcription.
4. During the emergent thematic coding reading analysis process, I wrote marginal notes throughout each transcribed document from each discussion set that reflected emerging themes pertinent to the purpose of the study.
5. Open, or initial, codes were attached to words, lines, and segments of the text that identified distinct themes related to the purpose of the study.
6. Marginal notes for each document set were grouped into a well-organized set of categories.
7. Examples of categories emerged based upon answers provided from focus group questions and included initial codes such as quality of performance assessments, preparedness of teachers to implement performance assessments, performance

assessments used to assess student learning, and appropriate and reliable outcomes of performance assessments.

8. Particular attention was given to information considered an outlier, or a perspective counter to the views of the most common responses as a tenet of Stake's Responsive program evaluation process which is "particularly alert to episodes that, however unrepresentative, add to understanding the complexity of the evaluand" (Stake, 2004, p. 88).
9. Each developed set of categories was contrasted and synthesized into a new coherent and comprehensive group faithful to the original set (Stufflebeam & Shinkfield, 2007).
10. Next, the comprehensive set was contrasted with the research questions guiding the study to develop a standardized set of categories known as axial codes. Axial, or focused, codes were applied to the continued reading of subsequent document sets until a relevant categorical code accompanied noteworthy segments of each document.

Summaries

For each discussion set, a summary was written in relation to each category of findings (Stufflebeam & Shinkfield, 2007). The summaries were identified with an *s,* for summary, and the coded letters and numbers that denoted the original relevant document, s-fg-1, s-fg-2, and so on. After looking across and examining the summaries for the different sets of categorical information, conclusions were written in relation to each research question. This practice helped guide the structure and description of the final report (Stufflebeam & Shinkfield, 2007).

Ethical Considerations

I requested permission from the Institutional Review Board (IRB) at William and Mary School of Education to use the collected data sources for the purpose of this study, as well as PCPS. To protect privacy, the names of participants, schools, and the school division were not disclosed in this study. Before any information is utilized in writing the final report of this study, all participants received an Informed Consent Agreement (Appendix H) that explained the purpose of the research study, the duration, what the study involved, and that the findings were to be made public. I explained that the study was completely voluntary and that participants were able to withdraw from the study at any time without penalty. I promised to protect participants and to ensure complete confidentiality by using pseudonyms in any written report of the focus group and in the final dissertation. I explained the benefits of the study and that there were no foreseeable risks in participating. As the curriculum specialist and researcher in this study, I did not use the influence of my position to compel teachers to participate in the research of this study. To mitigate biases, I maintained a neutral stance when participants gave answers to open-ended questions in order to avoid implying there was a right answer. I carefully followed a scripted discussion during each focus group about the purpose, permission to record conversations, and the option for participants not to partake without fear of retribution. "Complete descriptions of findings, limitations, and any resulting conclusions were made available to stakeholders, upon request" (Mertens & Wilson, 2012, p. 25).

Limitations, Delimitations, and Assumptions

The limitations of this study included the following:

1. The QCRT is presumed an adequate instrument to characterize performance assessment quality based on the fact the set of expectations was developed by the

VDOE for use statewide.

2. The number of participants in this study was contingent upon those who volunteered to participate.
3. The accuracy of the data collected about feelings, beliefs, and perceptions regarding performance-based assessments were contingent upon teacher-participants honestly sharing this information with me.
4. The recommendations made to strengthen performance assessments were limited to the perceptions of the sample population under study.
5. The QCRT findings were limited to the interpretation of two curriculum specialists in one school district.

Delimitations of the study included the following:

1. The participants are individuals who taught USI and USII courses and not other courses with local alternative assessments.
2. While other methods may have existed to reveal the quality of performance assessments, the Virginia QCRT for Performance Assessments was the only one used for the purpose of this study.
3. The study examined performance assessments that were not created or critiqued by me using Virginia QCRT for Performance Assessments before being implemented with students.
4. The teacher selected performance assessments submitted for the purpose of this study were not representative of the entire population of teachers who are required to implement alternative assessments.
5. Neither student perceptions nor achievement data was considered in this study.

6. Even though individual participants were used to identify categories in the focus group data, the analysis presented was representative of the school. For example, if Participant A from School A mentioned Pinterest as their answer, then Pinterest became part of the School A data.

Certain assumptions were made in this research study and should be noted:

1. The QCRT was an effective instrument to review the quality of performance assessments.
2. The data collected in this study would be used to identify and develop future quality performance assessments.

CHAPTER 4

FINDINGS

The purpose of this study was to find characteristic features of quality in performance-based assessments through a responsive evaluation approach. To address the research questions, a collection of artifacts was examined, as well as data obtained from focus groups. In total, 24 teacher-selected social studies performance assessments, two from each of the 12 middle schools, were reviewed for quality using the Virginia Quality Criteria Review Tool (QCRT) for Performance Assessments. Focus group data were collected from USI and USII teachers in Performance County Public Schools. While PCPS has 12 middle schools and a combined total of 76 USI and USII teachers, not all schools and teachers participated in focus groups. Focus group data were collected from 10 middle schools. There is a combined total of 58 USI and USII teachers in the 10 schools. From the 10 participating schools, 29 teachers volunteered to contribute to focus group discussions.

Thus, 83% of middle schools in PCPS contributed to focus groups with a 45% participation rate among teachers from 10 middle schools. Even though individual participants were used to identify categories, the analysis presented is representative of the school. For example, if a participant from a school mentioned Pinterest as an answer, then Pinterest becomes part of the school data. Finally, data were analyzed to determine to what extent a sample group of schools were able to recognize quality characteristics in performance assessments. Table 8 presents an overview of the data sources available for analysis in this study.

Table 8

Overview of Data Sources

Middle School	Combined USI and USII Performance Assessments	Focus Group Participants
1	2	4
2	2	2
3	2	0
4	2	3
5	2	3
6	2	2
7	2	2
8	2	0
9	2	3
10	2	3
11	2	3
12	2	4
Total	**24**	**29**

Note. United States History to 1865 (USI) and United States History 1865 to the present (USII)

The results of the analysis presented in this chapter answer each of the four research questions. Data collected from the QCRT were used to determine to what degree performance assessments, used as local alternative assessments to the Virginia Standards of Learning (SOL) tests for accountability in PCPS, met the state's criteria for quality. QCRT data were clearly identified and tagged as originating from courses *United States History to 1865 (USI)* and *United States History 1865 to the Present (USII)* along with school designation*(s).* An example of this

designation is “USI.S1.”

Next, data gleaned from focus groups revealed how performance-based assessments were being implemented in two middle school social studies courses, perspectives on the merit and worth of said performance assessments and recommendations for developing and using quality performance assessments moving forward. Any references made to teachers or schools within the data were replaced with a pseudonym. In this analysis, distinctions were made and noted between USI and USII performance assessments. Alternatively, focus group data were reviewed and analyzed by each school collectively and not distinguished by course. As well, a description of the PCPS alternative assessment structure is provided to better understand the context in which performance assessments are implemented.

The performance-based assessments implemented in the USI and USII courses fit within a structural design known as the *PCPS Alternative Portfolio Assessment* (Appendices C and D). The alternative portfolio assessment most resembles a showcase portfolio approach to assessment while incorporating features of growth and evaluation as well as a “snapshot” of student work. Students choose between two to four artifacts that represent their best work for the various topics taught and assessed during each 9-week grading period. Students conduct self-evaluations by rating their progress on a four-point scale and meet in small groups to showcase their work. Teacher-selected performance-based assessments were among the artifacts from which students could choose to put into their portfolio and are the subject of a quality review in this current study. For the purposes of this study, *teacher-selected* performance-based assessments can be described as assessments created or selected by teachers that best represent the proficiency of the knowledge and skills students need to attain related to topics under study.

Data Disaggregation

QCRT data were disaggregated by school, course and criterion, as seen in Tables 11-15 and answer Research Question 1. Tables include each school's QCRT quality ratings represented by percentages. Based on the fact that two raters reviewed each performance assessment for quality, a statistical procedure known as *percentage of agreement* was used to establish interrater reliability. To establish interrater reliability in this study, the two quality review team participants independently rated the quality of the performance assessments using the QCRT. Then, the quality review team engaged in the practice known as *comparative consensus* by stating their individual ratings for each criterion and noting the cited evidence necessary to support the ratings. Ratings that did not agree were deliberated until there was a consensus rating of 100% agreement.

To determine quality percentages per course and criterion, ratings for each criterion were added up and divided by the total number of points available. For example, if criterion ratings for a course added up to be 13 and the total number of points available were 17, then 13 would be divided by 17 and converted to a 76% quality rating, as shown here:

1. Count the number of ratings in agreement (e.g., 13).
2. Count the total number of ratings (e.g., 17).
3. Divide the number of ratings in agreement by the total number of ratings to get a fraction (e.g., 13/17).
4. Convert to a percentage (e.g., 13/17 = 76%).

In addition, QCRT rating numbers are defined as follows: 0 = *No Evidence*; 1= *Limited Evidence*; 2= *Partial Evidence*; and 3= *Full Evidence*. Evidence to substantiate Criterion 1 data was communicated in written summaries by the two raters to justify their scores on the original

QCRT documents (see Appendix H for an example justification).

Summaries addressed the accuracies and deficits found in the performance assessments. The breakdown of QCRT data per sub-criterion and by school identified scaled score numerical data based on the rubric for quality ratings, as well as a USI and USII combined quality rating. All sub-criteria were scored separately, and an overall average per sub-criterion was calculated. The overall quality rating for each criterion was determined by finding the average of the quality ratings by course or by finding the average of the sub-criterion quality ratings.

Data disaggregation for Research Questions 2, 3, and 4 were noted later in Chapter Four. Correlation was made between said research questions and answers to focus group questions to create findings that addressed the perceptions of USI and USII teachers regarding the PCPS process of implementing performance-based assessments. Descriptive words and phrases that emerged throughout the data collection were disaggregated and presented in Tables 18, 19, and 20.

Findings for Research Question 1: **To what degree do performance assessments used as local alternative assessments to Virginia Standards of Learning tests for accountability meet the state's criteria for quality?**

In order to answer Research Question 1, a collection of 24 performance assessments were reviewed for quality using the QCRT. The QCRT had seven criteria with numerical rubrics that presented quality ratings as follows: 0 = *No Evidence*; 1=*Limited Evidence*; 2 = *Partial Evidence*; and 3 = *Full Evidence* (Appendix B). Quality review raters compared every performance assessment to the seven criteria and assigned quality ratings. Interrater reliability was established through comparative consensus until there were agreement ratings of 100%. Tables 11-15 reflect the criteria rating scores. After each table, a narrative and graph(s)

summarize the data to explicate findings.

Criterion 1: Standards/Intended Learning Outcomes

The first criterion is intended to gauge and reveal the following quality characteristics of alignment to the Virginia SOL and intended learning outcomes, as shown in Table 9. There are three sub-criteria for Criterion 1. Sub-criterion A referred to the alignment of SOL content topics. Sub-criterion B represented opportunities for application of disciplinary or cross-disciplinary concepts and for transferable skills such as application, analysis, evaluation, synthesis, or original creation. Sub-criterion C represents opportunities for students to develop and demonstrate disciplinary or cross-disciplinary deeper learning competencies such as how to think critically, problem-solve, communicate effectively, work collaboratively, and effectively utilize technology.

Criterion 1 Summary. A summary of Criterion 1 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative. Table 9 showed data from Sub-criteria A, B and C of Criterion 1 and an overall quality average by school and district. Next, a visual display of data was presented by way of bar graphs and respective narrative descriptions. Bar graphs provided an opportunity to see quality ratings by school and a comparison of data among schools in the district.

Table 9

Criterion 1 Data by School and District

	Sub-Criteria			
School	**A**	**B**	**C**	***M* Quality Rating**
1	2	1.5	1	1.5
2	.5	2.5	1.5	1.5
3	1	2	2	1.7
4	.5	.5	.5	.5
5	2	2.5	2.5	2.3
6	3.	2.5	2	2.5
7	1.5	2	1.5	1.7
8	0	2	1.5	1.2
9	0	3	1.5	1.5
10	0	1.5	1.5	1
11	2.5	3	3	2.8
12	1.5	3	2.5	2.3
Average Sub-criteria Ratings	1.2	2	1.8	PCPS Overall Quality Rating for Criterion 1 1.**7**

Note. 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence; PCPS = Performance County Public Schools

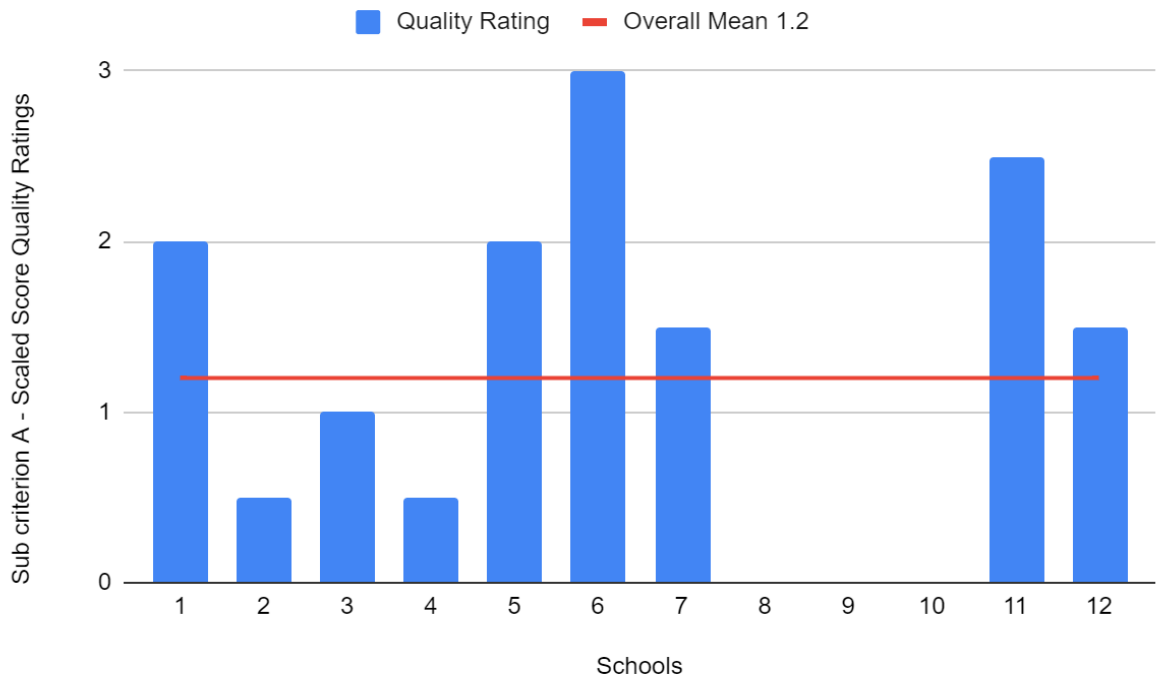
Next, bar graphs provided visual representations of sub-criteria data by school. Each bar was denoted by a number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school were listed on the vertical axis. Each school's bar on the graph depicted data reported in Appendix J. Bar graphs provided an opportunity to display data by school in order to make comparisons among schools in the district.

Figure 4 displayed data for Sub-criteria A and revealed quality ratings specifically related to the alignment of PAs to SOL and intended learning outcomes. School 6 data revealed full evidence (3) of alignment to Sub-criteria A in comparison to no evidence (0) in Schools 8, 9, and

10. While Schools 7 and 12 approached partial evidence (1.5) ratings, Schools 1 and 5 achieved partial evidence ratings (2). School 3 revealed limited evidence (2) ratings for Sub-criteria A, and Schools 2 and 4 only approached a limited rating (.5) of quality for Sub-criteria A. There was clear inconsistency of performance assessment alignment to content standards among middle schools in the district.

Figure 4

Criterion 1 Sub-Criterion A - Alignment to Content Standards



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

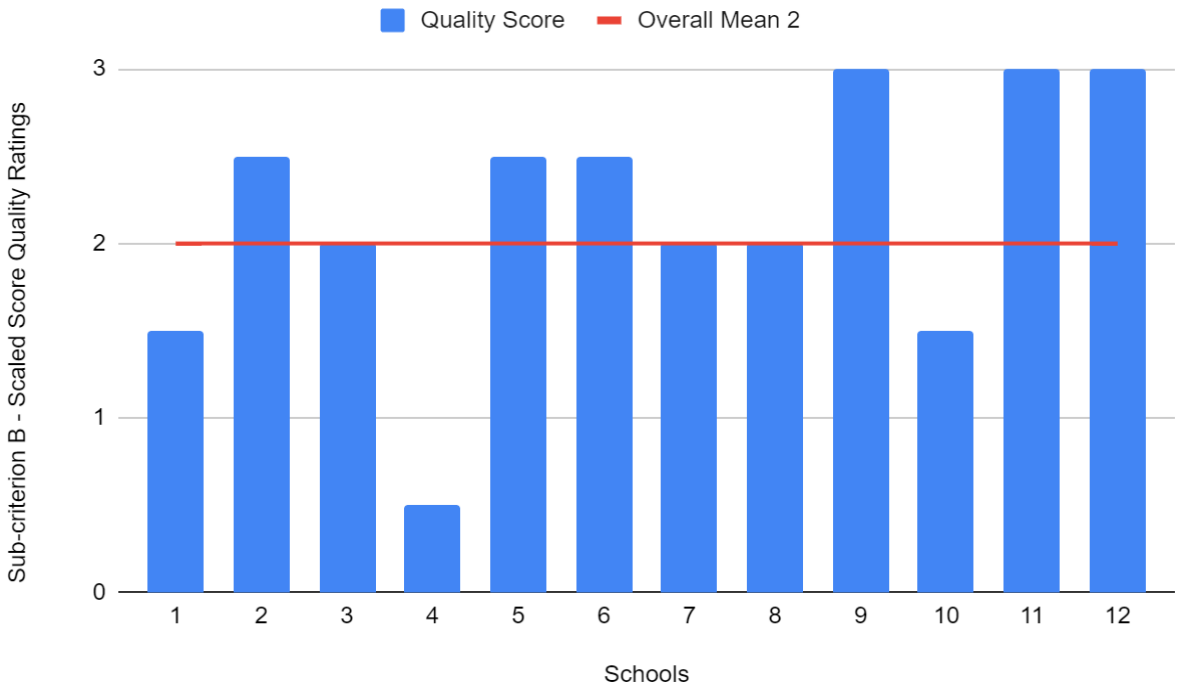
Figure 5 displays data for Sub-criterion B and reveals quality ratings specifically related to the alignment of PAs to the application of skills. Data for Schools 9, 11, and 12 revealed full evidence of quality (3), while Schools 2, 5, and 6 approached full evidence (2.5). Partial evidence (2) of quality was revealed in Schools 3, 7, and 8, while Schools 1 and 10 approached

partial evidence (1.5). School 4 data approached limited evidence (.5) of alignment to the application of skills and was an outlier of Sub-criterion B.

Skills alignment can add to a robust learning experience in performance assessments. Earlier in this study, Table 4 identified the frequency of common words and concepts associated with performance assessments by various researchers and revealed the highest number of occurrences as skills/abilities (Bland & Gareis, 2018). Worthy of note in the current study is the consistency of strong skills alignment among schools and the aforementioned corroborated evidence seen in Table 4.

Figure 5

Criterion 1 Sub-Criterion B - Skills Application.



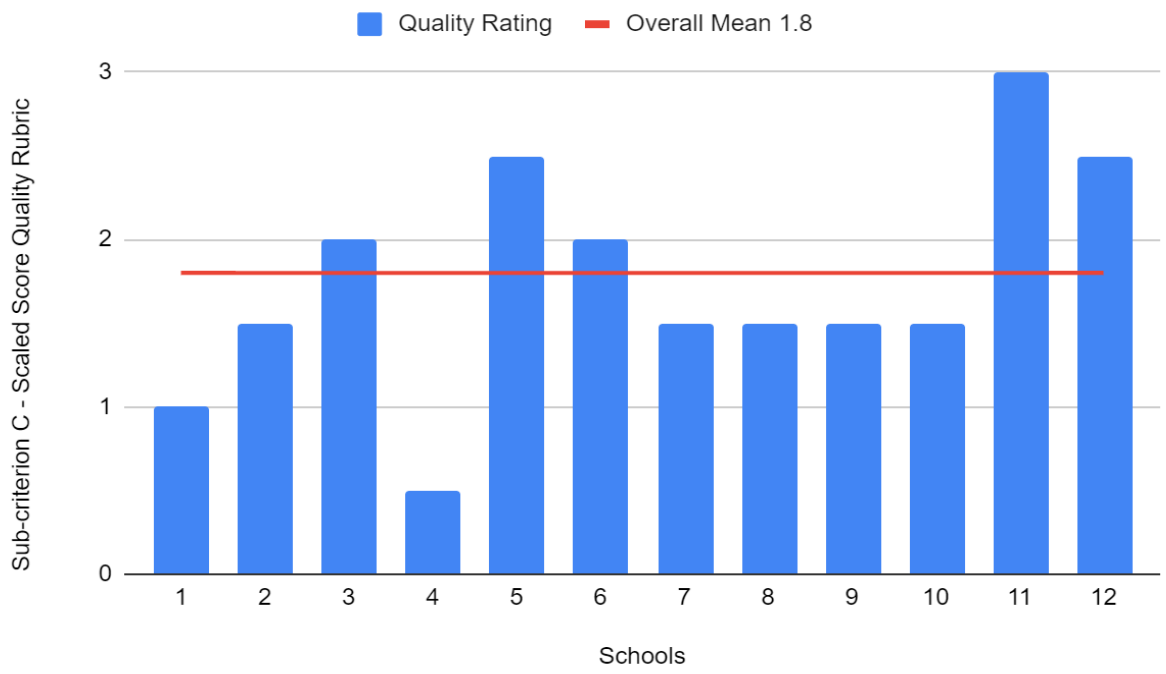
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 6 displayed data for Sub-criterion C and revealed quality ratings specifically

related to the alignment of PAs to deeper learning competencies. School 7 data displayed full evidence (3) of deeper learning competencies, while Schools 5 and 12 approached full evidence (2.5). Partial evidence was shown in the data for Schools 3 and 6, while Schools 2, 7, 8, 9, and 10 data approached partial evidence (1.5). School 1 and School 4 presented limited evidence (1). Worthy of note was the trend based upon seven schools that approached or achieved partial evidence. School 4 was an outlier with below limited evidence (.5).

Figure 6

Criterion 1 Sub-Criterion C - Deeper Learning Competencies



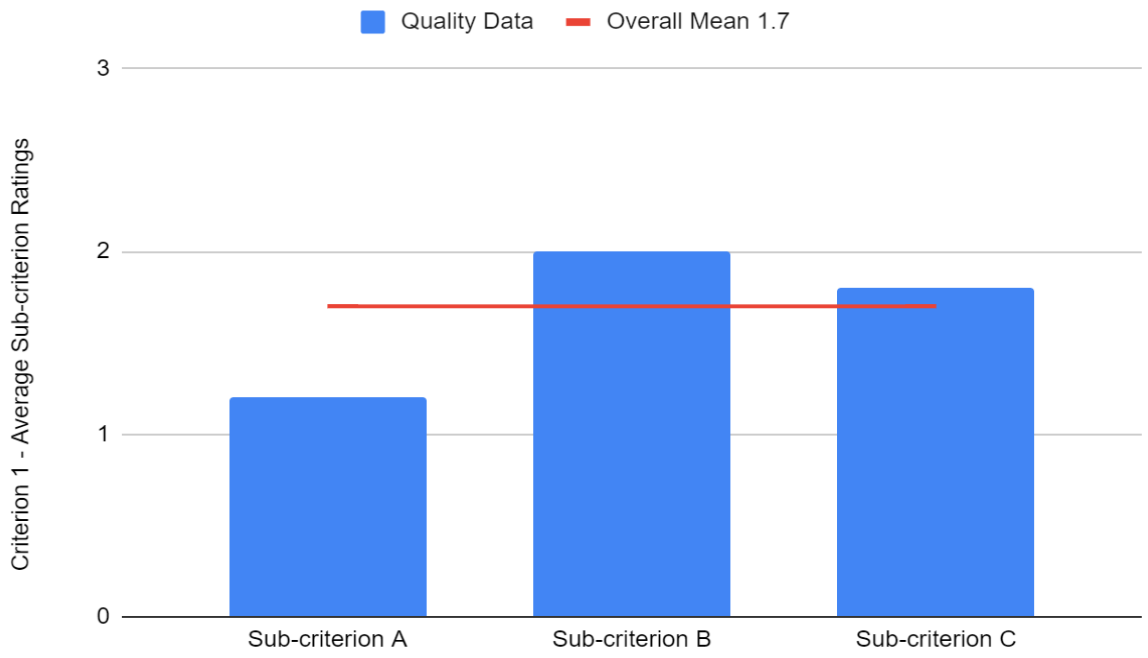
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 7 displayed the overall average ratings data for Sub-criteria A, B and C. Alignment to content standards (Sub-criterion A) was barely above limited evidence (1.2) in an approach to partial evidence. Alignment to skills application (Sub-criterion B) showed a rating of partial evidence (2), while deeper learning competencies (Sub-criterion C) was a near partial

evidence (1.8) rating. Overall, Criterion 1 data indicated weaknesses in the alignment of PAs to SOL and intended learning outcomes; School 4 consistently provided less than limited evidence of quality ratings in all three sub-criteria. All USI and USII teachers in PCPS would benefit from professional development in aligning PAs to SOL and intended learning outcomes.

Figure 7

Criterion 1 - Average Sub-Criterion Ratings



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

The next section revealed the second criterion's endeavor to rate and reveal the following characteristic of quality: Authenticity.

Criterion 2: Authenticity

Criterion 2 was introduced in Table 10 and focused on relevancy. What students are asked to do and for whom can create a scenario where the topic and context are meaningful and

relevant to the real-world. Real-world refers to students' communities, interests, or future careers. Real-world also referred to work authentic to the discipline, like evaluating historical sources as a historian would do. Findings indicated partial to full evidence (2.3) of PA alignment to Authenticity; thus, Criterion 2 was rated the highest among all quality criteria.

In the next section, a summary of Criterion 2 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative.

Criterion 2 Summary. Criterion 2 data was seen in Appendix K and presents a combined overall average by school and district. Next, a visual display of data was presented by way of a bar graph and narrative description. Bar graphs provided an opportunity to see quality ratings by school and a comparison of data among schools in the district.

Table 10

Criterion 2 Data by School & District

School	Quality Rating
1	1.0
2	3.0
3	2.0
4	1.5
5	3.0
6	3.0
7	2.5
8	2.0
9	2.5
10	1.5
11	3.0
12	3.0
Overall	**2.3**

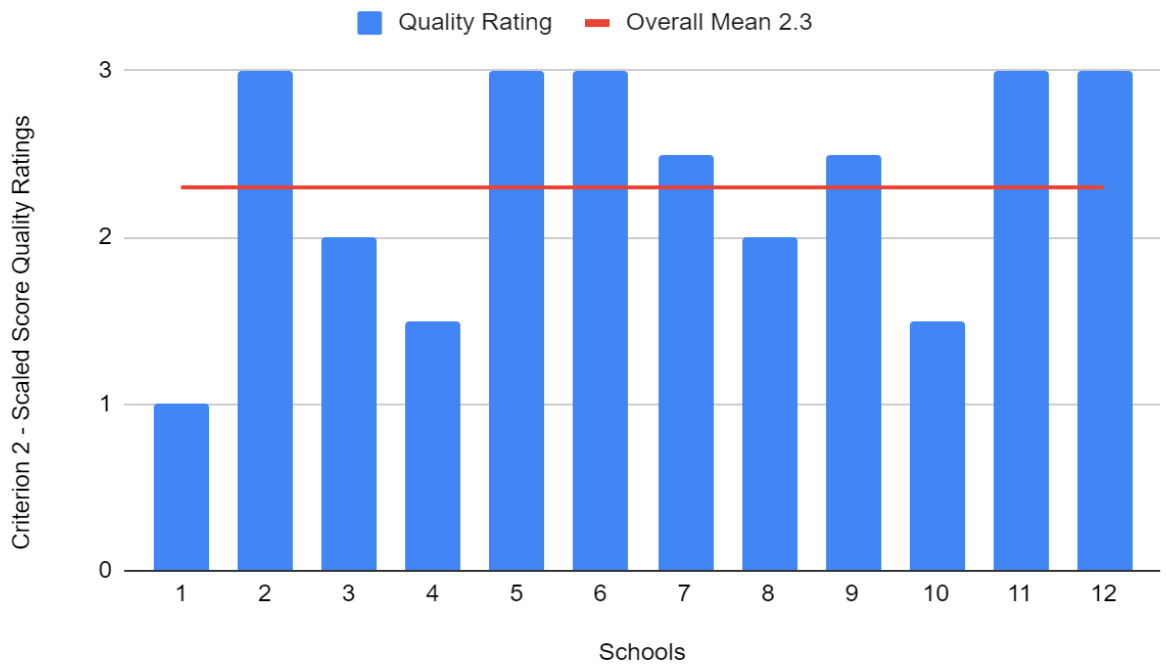
Bar graphs provide visual representations of data by school. Each bar is denoted by a

number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school are listed on the vertical axis. Each school's bar on the graph depicts data reported in Table 10.

Figure 8 displays data for sub-criterion A and reveals quality ratings specific to the alignment of PAs to Criterion 2. Data for Schools 2, 5, 6, 11, and 12 displayed full evidence (3) of alignment to Criterion 2, while Schools 7 and 9 approached full evidence (2.5). Schools 3 and 8 data displayed partial evidence (2), while Schools 4 and 10 approached partial evidence (1.5). School 1 data revealed limited evidence (1).

Figure 8

Criterion 2 – Authenticity



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

The next section revealed the third criterion's endeavor to rate and reveal the following characteristic of quality: language use for expressing reasoning.

Criterion 3: Language Use for Expressing Reasoning

Two sub-criteria are introduced in Appendix L. Sub-criterion A focused on providing multiple opportunities to access and use developmentally appropriate academic and disciplinary language to express reasoning. Sub-criterion B stated that various forms of language media, such as text, video, audio, or oral (i.e., conversations with peers), should be required as ways for students to use academic or disciplinary language to express their reasoning.

In the next section, a summary of Criterion 3 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative.

Criterion 3 Summary. A summary of Criterion 3 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs and narrative. Table 11 showed data from Sub-criteria A and B of Criterion 3. The table presents a combined, overall quality average by school, district quality ratings by sub-criterion, and an overall district quality rating for Criterion 3. Next, each sub-criterion displayed data by way of bar graphs. The visual representation of data in bar graphs provided an opportunity to see quality ratings by school and a comparison among schools in the district.

Table 11

Criterion 3 Data by School & District

Schools	Sub-criterion A	Sub-criterion B	Average Quality Ratings
1	2.5	2	2.3
2	3	3	3
3	3	2.5	2.8
4	1	1	1
5	2	2	2
6	3	1.5	2.3
7	2.5	2.5	2.5
8	1.5	1.5	1.5
9	2.5	1	1.8
10	1	1.5	1.3
11	3	3	3
12	3	2	2.5
Average Sub-criterion Ratings	2.3	2	Overall Quality Rating for Criterion 3: 2.2

Note. 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

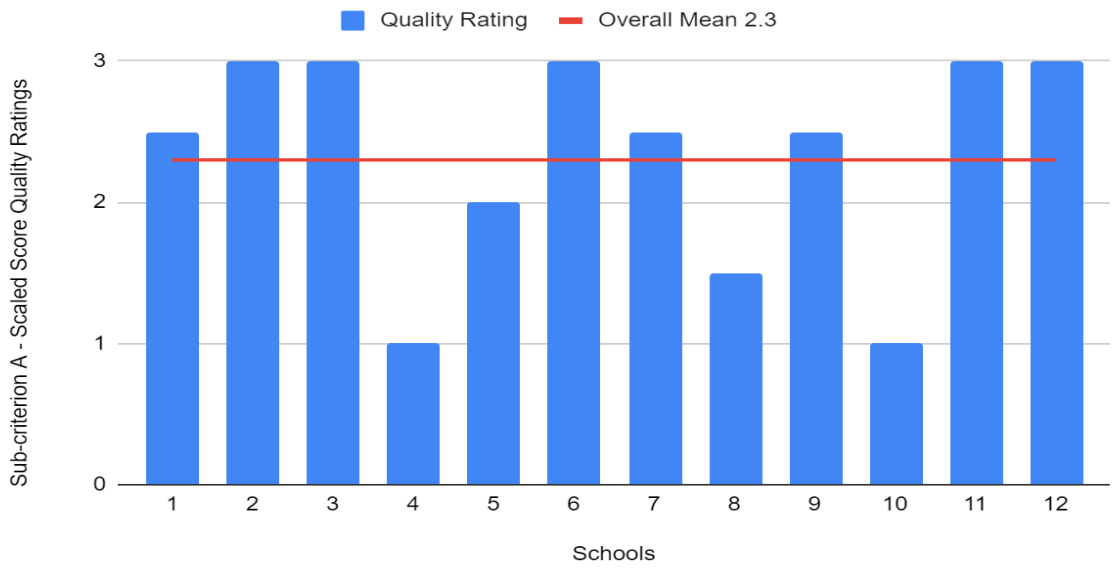
Next, bar graphs provided visual representations of sub-criterion data by school and among schools. Each bar was denoted by a number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school were listed on the vertical axis. Each school's bar on the graph depicted data reported in Table 11.

Figure 9 displayed data for Sub-criterion A and revealed quality ratings specifically related to the alignment of PAs to accessing and using developmentally appropriate academic and disciplinary language to express reasoning. Data for Schools 2, 3, 6, 11, and 12 revealed full evidence (3) of alignment to Sub-criterion A, while Schools 1, 7, and 9 approached full evidence (2.5). School 5 data showed partial evidence (2) of alignment, while School 8 approached partial alignment 1.5). Schools 4 and 10 data revealed limited evidence (1) of alignment to Criterion 3.

Worthy of note was that 8 out of 12 schools approached or attained full evidence of alignment to Sub-criterion A of Criterion 3.

Figure 9

Criterion 3 Sub-Criterion A - Access Appropriate Language

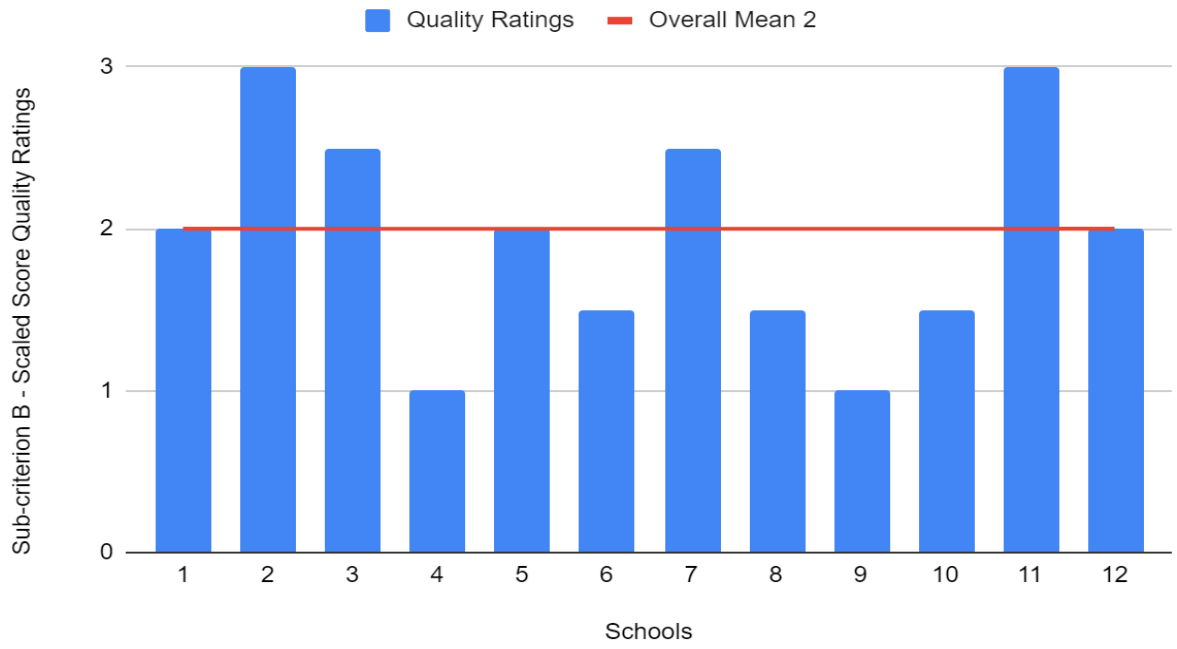


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 10 displayed data for Sub-criterion B and revealed quality ratings specifically related to the alignment of PAs to using one or more forms of language to express reasoning. Schools 2 and 11 data revealed full evidence (3) of alignment to Sub-criterion B, while Schools 3 and 7 approached full evidence (2.5). Schools 1, 5, and 12 data showed partial evidence (2) of alignment, while Schools 6, 8, and 10 approached partial alignment (1.5). Schools 4 and 9 data revealed limited evidence (1) of alignment to Criterion 3. Worthy of note was that 8 out of 12 schools approached or attained full evidence of alignment to Sub-criterion A of Criterion 3.

Figure 10

Criterion 3 Sub-Criterion B - Use One or More Forms of Language



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 11 displayed the overall average ratings data for Sub-criteria A and B. Alignment to accessing appropriate language (Sub-criterion A) was barely above a partial evidence (2.3) rating in an approach to full evidence. Alignment to using one or more types of appropriate language (Sub-criterion B) showed a rating of partial evidence (2). Overall, Criterion 3 data indicated some strengths in the alignment of PAs to language use for expressing reasoning

Figure 11

Criterion 3 Language Use for Expressing Reasoning by Sub-Criterion Average

Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

The next section will reveal the fourth criterion's endeavor to rate and reveal the following characteristic of quality: success criteria for students.

Criterion 4: Success Criteria for Students

Three sub-criteria are introduced in Appendix M. Sub-criterion A stated that a rubric or other scoring tools tightly aligned to the intended learning outcomes were included. Sub-criterion B stated that the scoring tool should be clear and concise and written in audience-friendly language so that students understand the expectations. As well, the tool should provide a space for feedback to students about their work and how it can be improved. Sub-criterion C states that a scoring tool should be used across performance assessments within a course to support a consistent set of expectations to students and observe student growth over time.

In the next section, a summary of Criterion 4 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative.

Criterion 4 Summary. Table 12 shows data from Sub-criteria A, B, and C of Criterion 4. The table data presented a combined, overall quality average by school, district quality ratings by sub-criterion, and an overall district quality rating for Criterion 4. Next, data for each sub-criterion presented is displayed of data by way of bar graphs. The visual representation of data in bar graphs provide an opportunity to see quality ratings by school and a comparison among schools in the district.

Table 12

Criterion 4 Data by School & District

Schools	Sub-criterion A	Sub-criterion B	Sub-criterion C	Average Quality Rating
1	1	0.0	0	0.3
2	0.5	1.5	1.0	1.0
3	0.5	0.5	0	0.3
4	1	1	0.5	0.8
5	0.5	0.5	0.5	0.5
6	3	2.5	1	2.0
7	2.0	2	2.5	2.2
8	1.5	1.5	0.0	1.0
9	1	1.0	0.0	0.7
10	0	0.0	0.0	0
11	1.5	1.0	1.0	1.2
12	0.5	0.5	0	0.3
Average Sub-criterion Ratings	1.0	1	0.5	PCPS Overall Quality Rating for Criterion 4 0.8

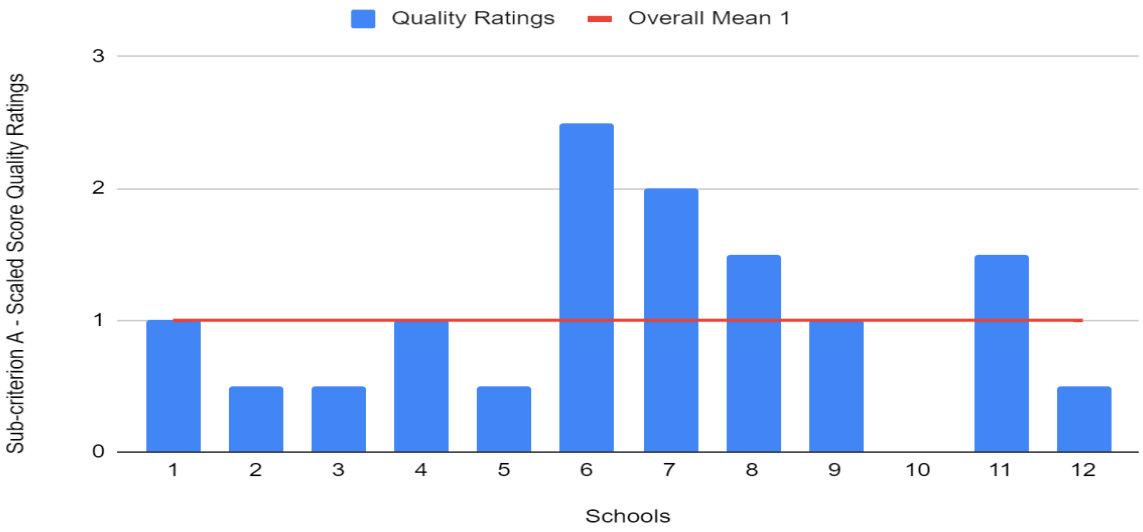
Note. 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence; PCPS = Performance County Public Schools.

Next, bar graphs provide visual representations of sub-criterion data by school and among schools. Each bar was denoted by a number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school were listed on the vertical axis. Each school's bar on the graph depicted data reported in Table 12.

Figure 12 displayed data for Sub-criterion A, which stated that a rubric or other scoring tools tightly aligned to the intended learning outcomes was included. School 6 data showed full evidence (3). School 7 data showed partial evidence (2) while Schools 8 and 11 approached partial evidence (1.5). School 1, 4 and 9 showed limited evidence (1) while schools 2, 3, 5, and 12 approached limited evidence (1). School 10 presented no evidence of alignment to Sub-criterion A.

Figure 12

Criterion 4 Sub-Criterion A Scaled Score Quality Ratings



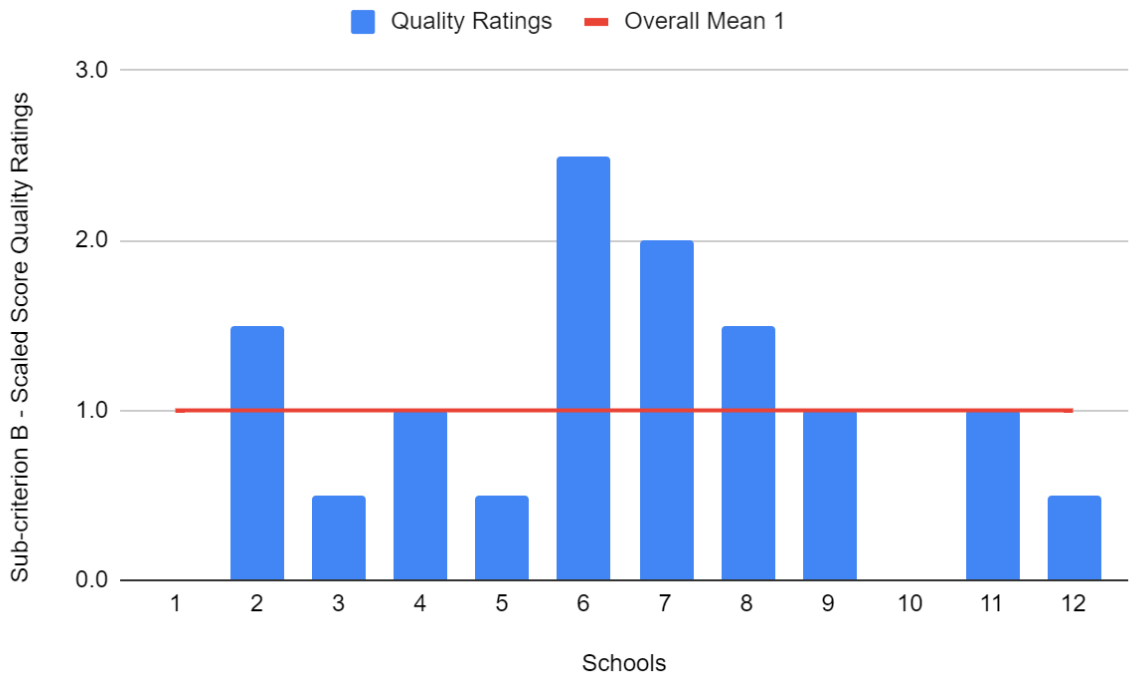
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 13 displays data for Sub-criterion B, which stated that the scoring tool should be clear and concise and written in audience-friendly language so that students understand the

expectations. As well, the tool should provide a space for feedback to students about their work and how it can be improved. School 6 data approached full evidence (2.5). School 7 showed partial evidence (2) while Schools 2 and 8 approached partial evidence (1.5). Schools 4, 9 and 11 data showed limited evidence (1) while Schools 3, 5 and 12 approached limited evidence (.5). No data (0) were provided from Schools 1 and 10 for sub-criterion B.

Figure 13

Criterion 4 Sub-Criterion B Scaled Score Quality Ratings

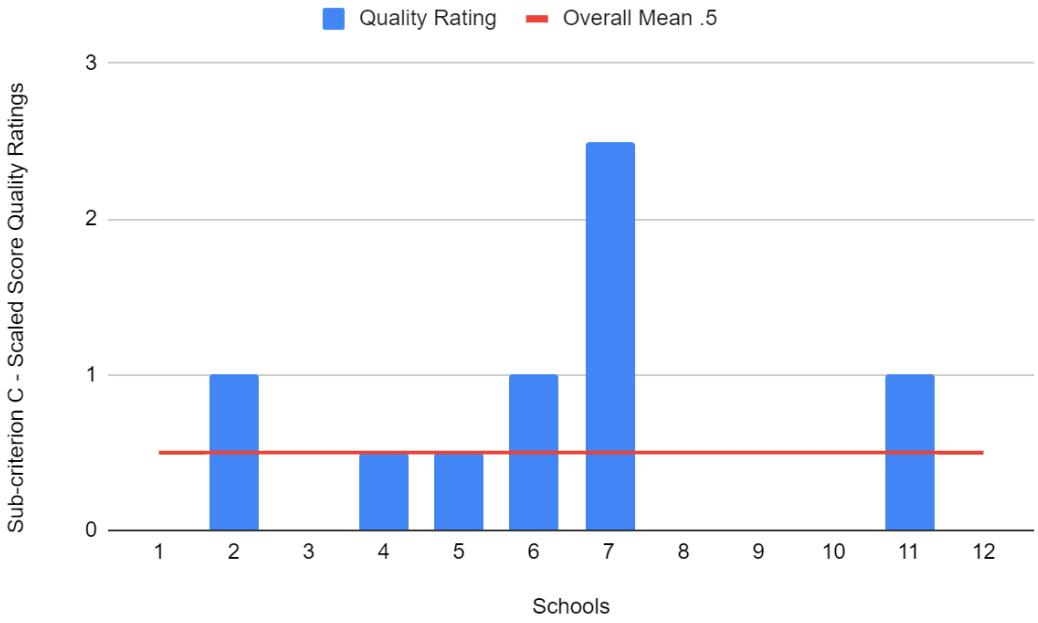


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 14 displays data for Sub-criterion C and shows that a scoring tool should be used across PAs within a course to support a consistent set of expectations to students and observe student growth over time. School 7 data approached full evidence (2.5). Schools 2, 6, and 11 revealed limited evidence (1). Schools 1, 3, 8, 9, 10, and 12 provided no data for Sub-criterion C (0). Sub-criterion C was a substantially weak area for most schools in the district.

Figure 14

Criterion 4 Sub-Criterion C Scaled Score Quality Ratings

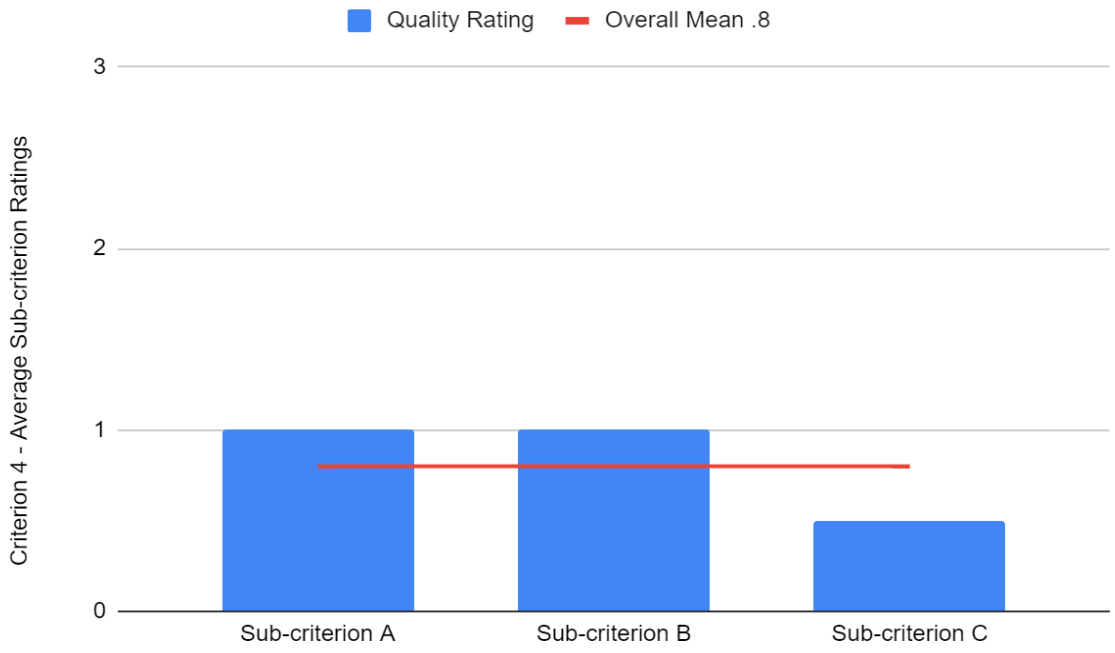


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 15 displays the overall average ratings data for Sub-criteria A, B and C. Alignment to tightly aligned rubrics (Sub-criterion A) was a rating of limited evidence (1). A clear and concise scoring tool written in audience-friendly language so that students understand the expectations and provide a space for feedback to students about their work and how it can be improved (Sub-criterion B) showed limited evidence (1) ratings. A scoring tool used across PAs within a course to support a consistent set of expectations to students and observe student growth over time (Sub-criterion C) revealed less than limited evidence (.5). Overall, Criterion 4 data indicated overall weaknesses in providing success criteria for students. All USI and USII teachers in PCPS would benefit from support in the development of success criteria for PAs.

Figure 15

Criterion 4 Average Sub-Criterion Ratings.



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

The next section will reveal the fifth criterion's endeavor to rate and reveal the following characteristics of quality: student directions, prompt and resources/materials.

Criterion 5: Student Directions, Prompt, and Resources/Materials

The three sub-criteria of criterion 5 are displayed in Appendix N. Sub-criterion A focused on alignment of student-facing prompts, directions, and resources to the intended learning outcomes. Sub-criterion B stated that the student-facing prompt, directions, and resources are clear, complete, written in accessible grade level language, and organized in an accessible format for students. Sub-criterion C referred to the scenario of the PA, as well as the materials and resources, being sensitive to the community and free of bias.

In the next section, a summary of Criterion 5 data presents the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative.

Criterion 5 Summary. A summary of Criterion 5 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative. Table 13 shows data from Sub-criteria A, B and C of Criterion 5. The table data presented a combined, overall quality average by school, and district quality ratings by sub-criterion, and an overall district quality rating for Criterion 5. Next, each sub-criterion presented a display of data by way of bar graphs. The visual representation of data in bar graphs provided an opportunity to see quality ratings by school and a comparison among schools and the district as a whole.

Table 13

Criterion 5 Data by School & District

School	Sub-criterion A	Sub-criterion B	Sub-criterion C	Average Quality Rating
1	0	1	1.5	.8
2	2	2	3.	2.3
3	1.5	1.5	3	1.7
4	1	1.5	1.5	1.3
5	1	2.5	3	2.2
6	2	2.5	2	2.2
7	2	2	3	2.3
8	2	2	1.5	1.8
9	.5	2	1.5	1.3
10	0	1	1.5	.8
11	2	3	3	2.7
12	1.5	1	3	1.8
Average Sub-criterion Ratings	1.3	2	1.8	PCPS Overall Quality Rating for Criterion 5 1.7

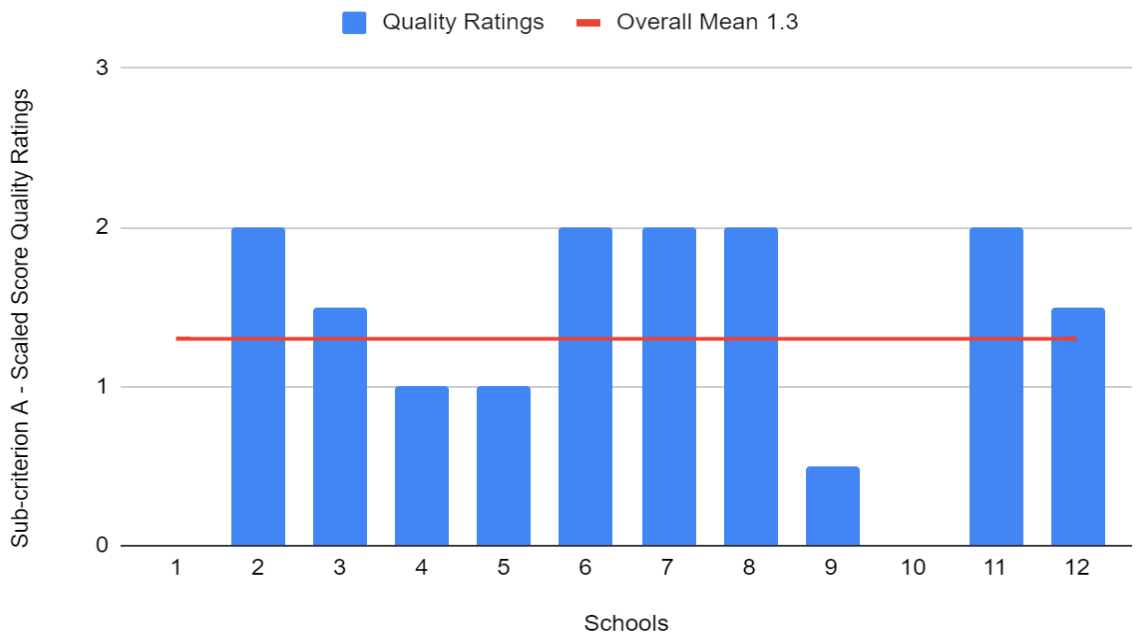
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence; PCPS = Performance County Public Schools.

Next, bar graphs provided visual representations of sub-criterion data by school and among schools. Each bar is denoted by a number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school are listed on the vertical axis. Each school's bar on the graph depicts data reported in Table 13.

Figure 16 displayed data for Sub-criterion A and focused on alignment of student-facing prompts, directions, and resources to the intended learning outcomes. Schools 2, 6, 7, 8, and 11 data revealed partial evidence (2), while Schools 3 and 12 approached partial evidence (1.5). Schools 4 and 5 show limited evidence (1) of alignment and School 9 approached limited evidence (.5). No evidence (0) of alignment was seen from Schools 1 and 10.

Figure 16

Criterion 5 Sub-Criterion A Scaled Score Quality Ratings

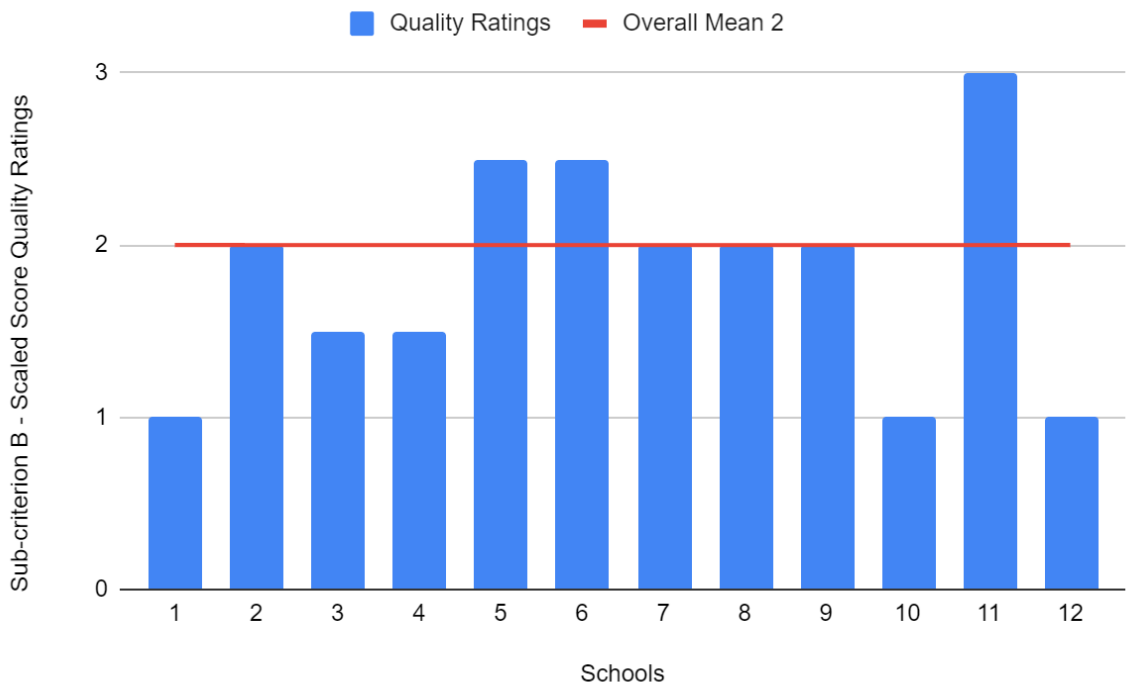


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 17 displays data for Sub-criterion B and shows that student-facing prompts, directions, and resources should be clear, complete, written in accessible grade level language and organized in an accessible format for students. School 11 data revealed full evidence (3) of alignment, while Schools 5 and 6 approached full evidence (2.5). Data from Schools 2, 7, 8, and 9 revealed partial evidence of alignment (2), while Schools 3 and 4 approached partial evidence (1.5). Schools 1, 10, and 12 data revealed limited evidence (1) of alignment to Sub-criterion B.

Figure 17

Criterion 5 Sub-Criterion B Scaled Score Quality Ratings



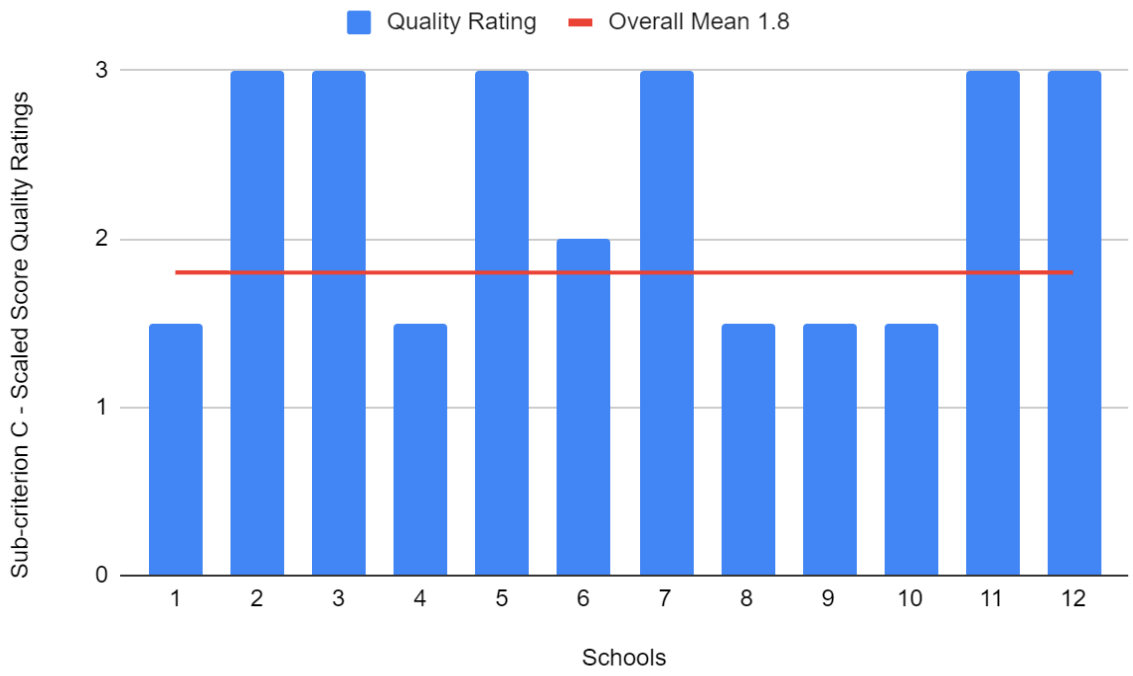
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 18 displayed data for Sub-criterion C and referred to the scenario of the performance assessment. A quality PA should have materials and resources that are sensitive to the community and free of bias. Schools 2, 3, 5, 7, 11, and 12 displayed full alignment (3) data.

School 6 data revealed partial alignment (2), while Schools 1, 4, 8, 9, and 10 approached partial (1.5) alignment.

Figure 18

Criterion 5 Sub-Criterion C Scaled Score Quality Ratings



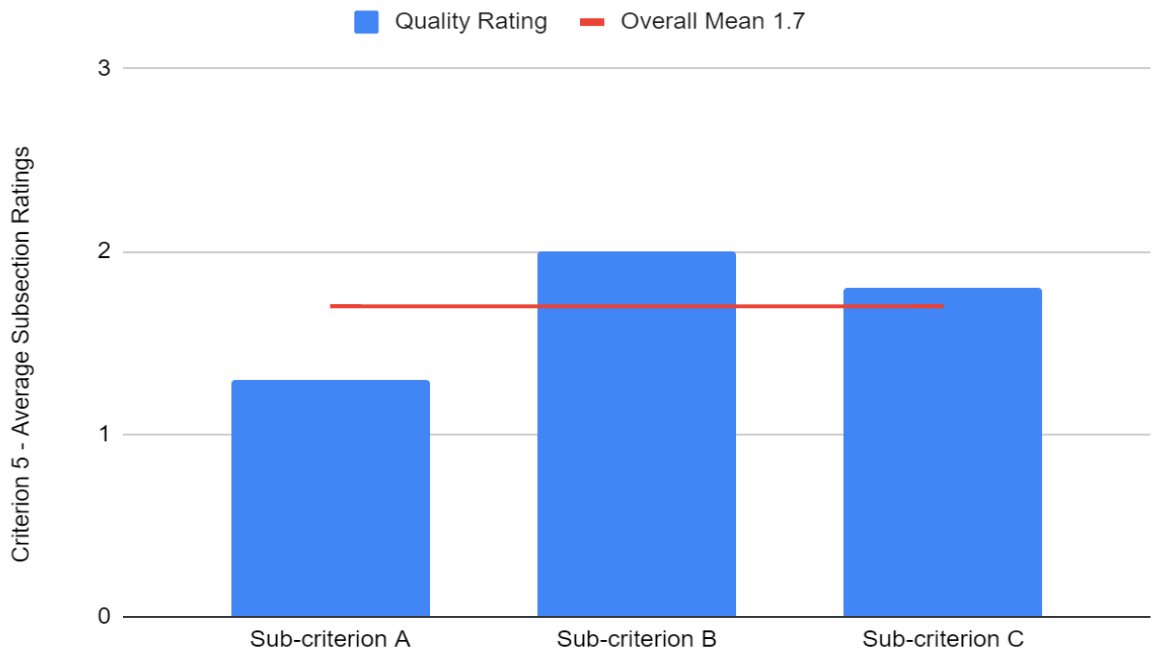
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidences

Figure 19 displayed Criterion 5's overall average ratings for Sub-criteria A, B, and C. Alignment of student-facing prompts, directions, and resources (Sub-criterion A) to the intended learning outcomes were barely above a rating of limited evidence (1.3). The clarity of student-facing prompts, directions, and resources (Sub-criterion B) received a rating of partial evidence (2). The PA scenario, materials, and resources being sensitive to the community and free of bias (Sub-criterion C) approached a partial evidence rating (1.8). Overall, while Criterion 5 data revealed partial to approaching partial evidence of QCRT alignment, all USI and USII teachers

in PCPS would benefit from support in the development of higher quality student directions, prompts, and resources/materials.

Figure 19

Criterion 5 Average Sub-Criterion Ratings



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

The next section reveals the sixth criterion's endeavor to rate and reveal the following characteristic of quality: accessibility.

Criterion 6: Accessibility

Two sub-criteria were part of Criterion 6 and are presented in Appendix O. Sub-criterion A focused on the design of the PA to accommodate the participation of all students. Teacher directions should include appropriate supports or alternatives to facilitate accessibility while maintaining the validity and reliability of the assessment. Sub-criterion B suggested options for accessibility be applied through the principles of Universal Design for Learning at the Center for

Applied Special Technology. Findings indicated limited evidence (1) of PA alignment to accessibility; Criterion 6 was the second lowest quality rating among all quality criteria.

In the next section, a summary of Criterion 6 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative.

Criterion 6 Summary. Table 14 showed data from Sub-criteria A and B of Criterion 6 as a combined quality average by school, district quality ratings by sub-criterion, and an overall district quality rating for Criterion 6. Next, each sub-criterion presented a display of data by way of bar graphs. The visual representation of data in bar graphs provided an opportunity to see quality ratings by school and a comparison among schools and the district as a whole.

Table 14

Criterion 6 Data by School & District

Schools	Sub-criterion A	Sub-criterion B	Average Quality Rating
1	0	1	0.5
2	0	.5	0.3
3	.5	1.5	1.0
4	0	0	0.0
5	1	.3	0.7
6	1.5	0	0.8
7	1.5	1	1.3
8	.5	1.5	1.
9	1	1.5	1.3
10	0	2.0	1.0
11	1.5	3.0	2.3
12	1	1.5	1.3
Average Sub-criterion Ratings	.7	1.2	PCPS Overall Quality Rating for Criterion 6:1

Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence; Performance County Public Schools (PCPS).

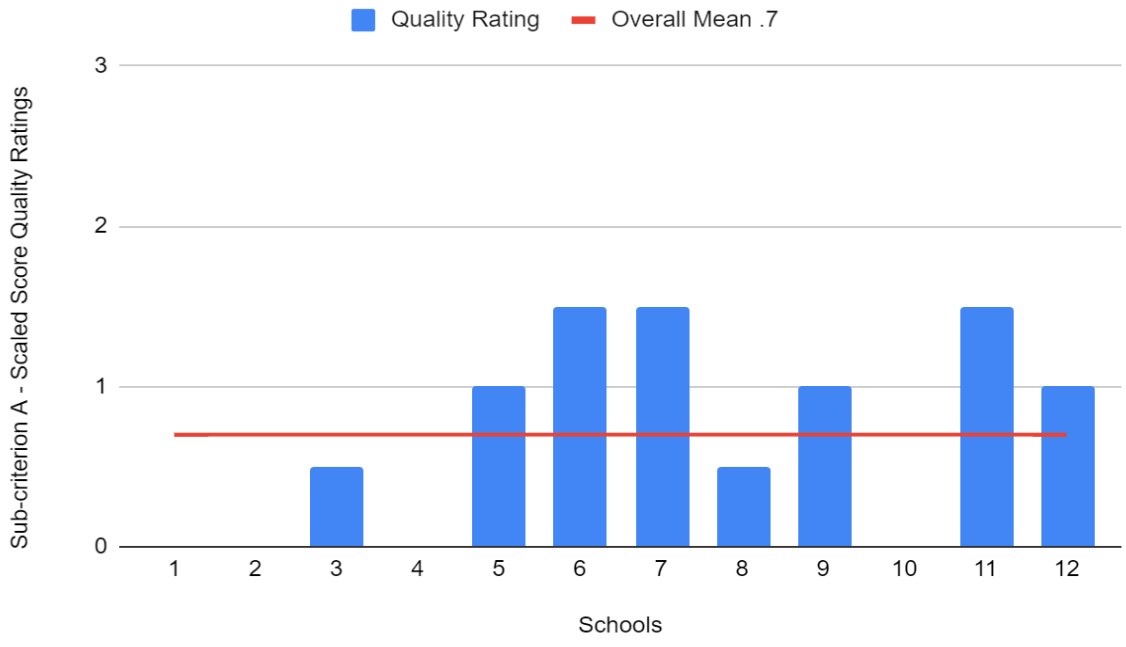
Next, bar graphs provided visual representations of sub-criterion data by school. Each bar

was denoted by a number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school were listed on the vertical axis. Each school's bar on the graph depicted data reported in Table 14. Bar graphs provided an opportunity to display data by school in order to make comparisons among schools.

Figure 20 displayed data for Sub-criterion A and focused on the design of the PA to accommodate the participation of all students. Schools 6, 7, and 11 revealed data approaching partial evidence (1.5). Schools 5, 9, and 12 showed limited evidence (1) of alignment, while Schools 3 and 8 approached limited evidence (.5). No evidence (0) of alignment was revealed from Schools 1, 2, 4, and 10. Sub-criterion A revealed very weak skills in the design of PAs to accommodate the participation of all students. Professional development for teachers to improve this skill is needed.

Figure 20

Criterion 6 Sub-criterion A Scaled Score Quality Ratings

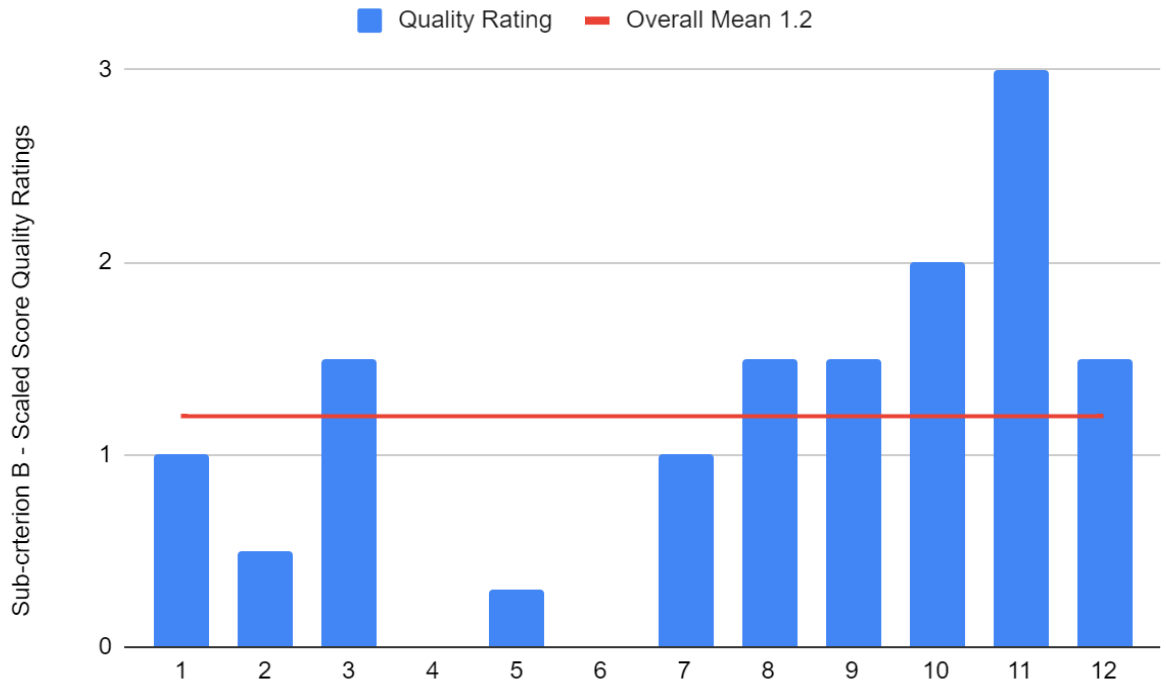


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 21 displayed data for Sub-criterion B and focused on providing options for differentiating the ways that students demonstrate their knowledge similar to the application of principles of Universal Design for Learning. School 11 provided full evidence (3) of alignment to Sub-criterion B and is an outlier of this data set. School 10 showed partial evidence (2.), while Schools 3, 8, 9, and 12 approached partial evidence (1.5). Schools 1 and 7 displayed limited evidence (1) of alignment, while Schools 2 and 5 approached limited evidence (.5 & .3, respectively). No evidence of alignment (0) was available from Schools 4 and 6. Overall, Sub-criterion B of Criterion 6 showed very weak alignment to differentiating the ways students are able to demonstrate their knowledge. As an outlier displaying full evidence, School 11 will be able to provide ideas about differentiation from which the other schools can learn.

Figure 21

Criterion 6 Sub-criterion B Scaled Score Quality Ratings

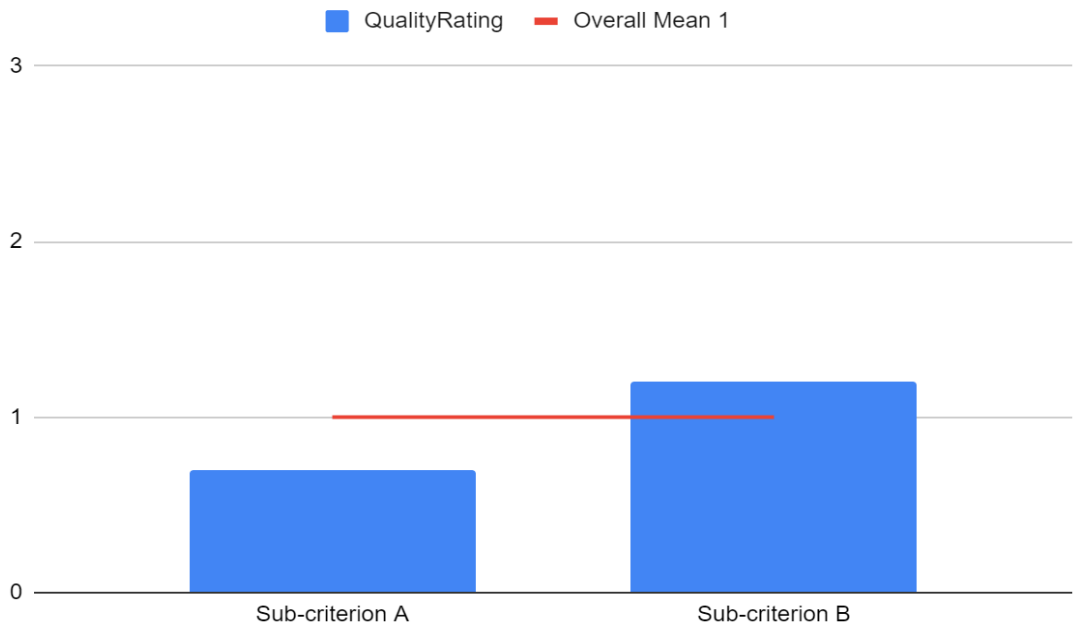


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 22 displayed Criterion 6's overall average ratings for Sub-criteria A and B. Alignment of the PA design to accommodate the participation of all students, as reflected in the teacher directions (Sub-criterion A), showed less than limited evidence. In most cases, teacher directions were not included. Options for differentiation (Sub-criterion B) showed data barely above limited evidence.

Figure 22

Criterion 6 Average Sub-criterion Ratings



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

The next section revealed the seventh and final criterion's endeavor to rate and reveal the following characteristic of quality: feasibility.

Criterion 7: Feasibility

Three sub-criteria to Criterion 7 were introduced and are presented in Appendix P. Sub-criterion A focused on the accessibility of resources and materials needed for the PAs and the

inclusion of student-facing prompts, directions, and scoring tools. Sub-criterion B looked at the complexity of the PA and whether the duration of implementation was realistic. Sub-criterion C referred to the inclusion of a schedule indicating how the PA is implemented across lessons and how the prior knowledge of students fits within a learning sequence.

In the next section, a summary of Criterion 7 data presented the average quality ratings by school and district in a variety of formats: table, bar graphs, and narrative.

Criterion 7 Summary. Table 15 showed data from Sub-criteria A, B and C of Criterion 7. The table presented a combined, overall quality average by school, district quality ratings by sub-criterion and an overall district quality rating for Criterion 7. Next, each sub-criterion presented a display of data by way of bar graphs. The visual representation of data in bar graphs provided an opportunity to see quality ratings by school and a comparison among schools and the district.

Table 15

Criterion 7 Data by School & District

Schools	Sub-criterion A	Sub-criterion B	Sub-criterion C	Average Quality Rating
1	1	0	0	.3
2	1	.5	0	.5
3	1.5	1	0	.8
4	1.5	0	0	.5
5	2	1.5	1.5	1.7
6	2	0	0	.7
7	3	3	1.5	2.5
8	2	1.5	0	1.2
9	1.5	0	0	.5
10	1	.5	.5	.7
11	2	0	0	.7
12	1	1	1	1
Average sub-criterion Ratings	1.6	.8	.4	PCPS Overall Quality Rating for Criterion 7: 0.9

Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

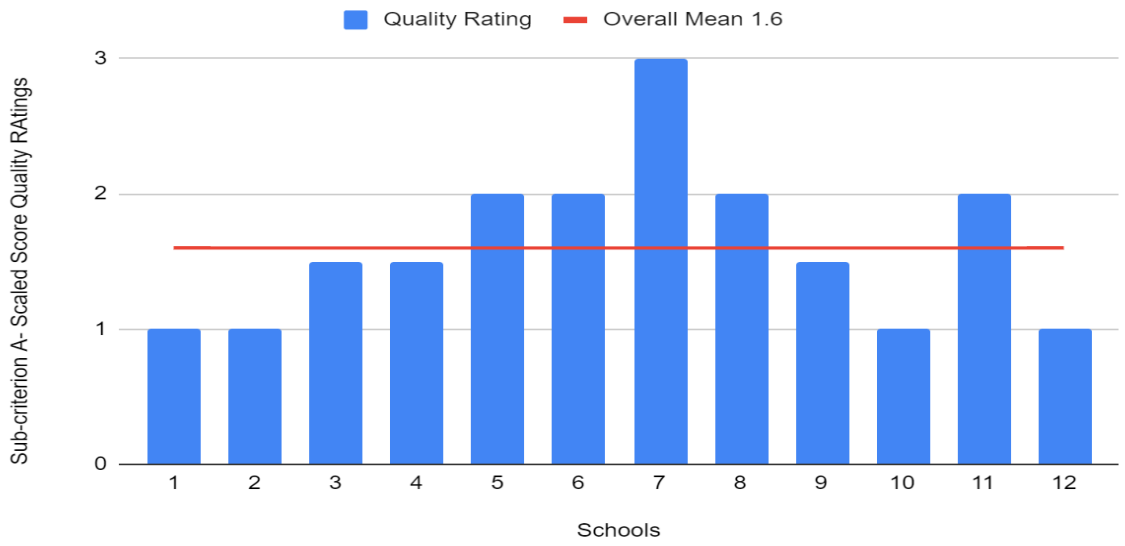
Next, bar graphs provided visual representations of sub-criterion data by school. Each bar is denoted by a number to represent each school, as seen on the horizontal axis. Scaled score quality ratings for each school were listed on the vertical axis. Each school's bar on the graph depicted data reported in Table 15. Bar graphs provided an opportunity to display data by school in order to make comparisons among schools.

Figure 23 displayed data for Sub-criterion A and focused on the accessibility of resources and materials needed for the PAs and the inclusion of student-facing prompts, directions, and scoring tools. School 7 data displayed full evidence (3) of alignment to Sub-criterion A. Schools 5, 6, 8, and 11 showed partial evidence (2), while Schools 3, 4, and 9 approach partial evidence (1.5). Schools 1, 2, 10, and 12 data revealed limited evidence (1) of alignment to Sub-criterion A.

Overall, Sub-criterion A data demonstrated strength in alignment to the accessibility of resources and materials needed to implement a PA, as well as, the inclusion of student-facing prompts, directions, and scoring tools.

Figure 23

Criterion 7 Sub-Criterion A Scaled Score Quality Ratings



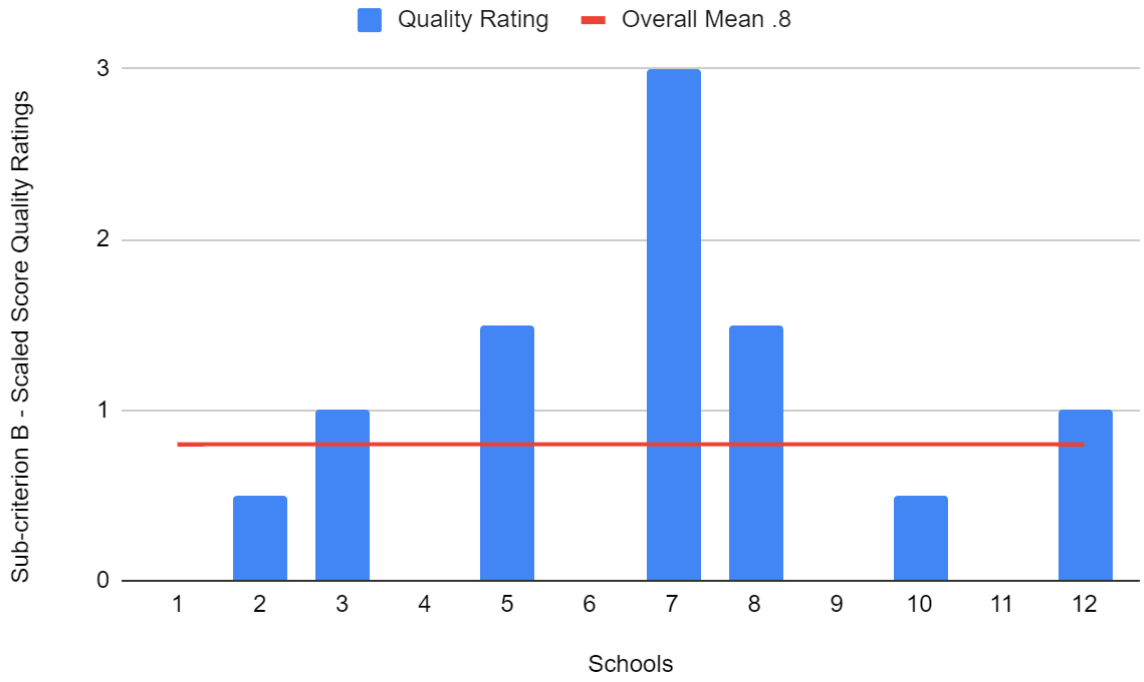
Note. Scale 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Next, Figure 24 displayed data for Sub-criterion B and revealed quality ratings specifically related to the complexity of the PA and whether the duration of implementation is realistic. School 7 data displayed full evidence (3) of alignment to Sub-criterion B and was a clear outlier because the next two closest schools, 5 and 8, merely approached partial evidence (1.5). Schools 3 and 12 revealed limited evidence (1), while schools 2 and 10 approached limited evidence (.5). Schools 1, 4, 6, 9, and 11 produced no evidence (0) of alignment to Sub-criterion B. Overall, there was very little evidence to show in Sub-criterion B that the duration of implementation was addressed based on the complexity of the PA. School 7's alignment to Sub-

criterion B is a strong example for other schools to consider in order to increase the quality of performance assessments.

Figure 24

Criterion 7 Sub-Criterion B Scaled Score Quality Ratings



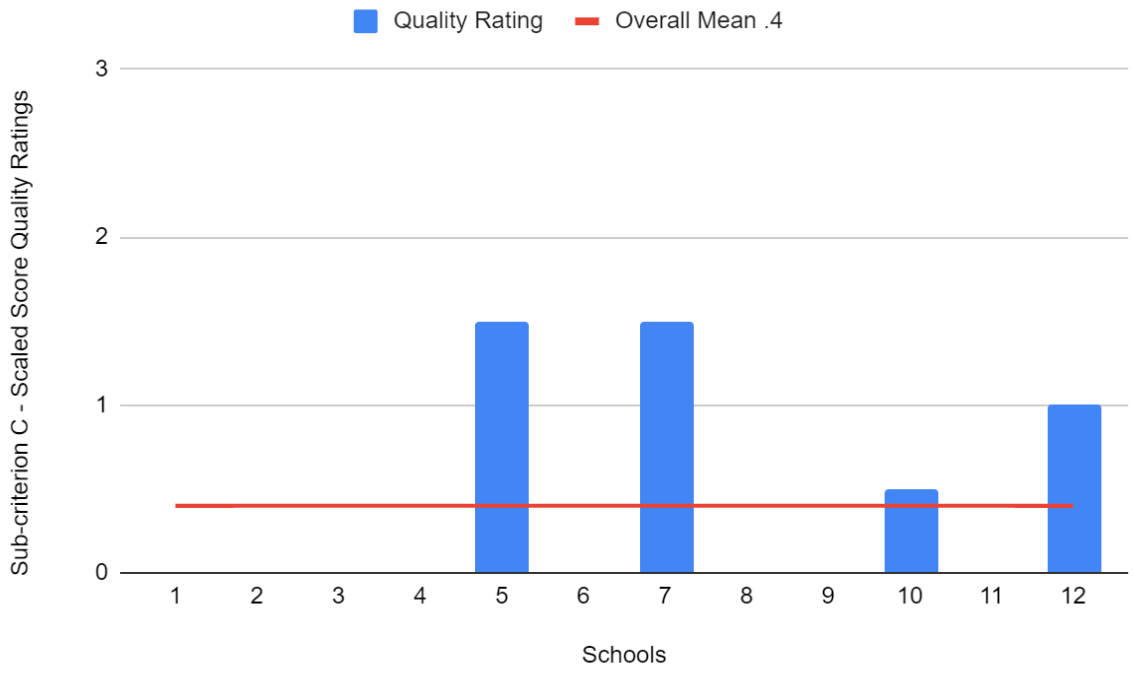
Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Finally, Figure 25 displayed data for Sub-criterion C and noted the inclusion of a schedule indicating how the PA was implemented across lessons and how the prior knowledge of students fit within a learning sequence. Schools 5 and 7 data approached partial evidence (1.5). School 12 displayed limited evidence (1), while School 10 approached limited evidence (.5). Schools 1, 2, 3, 4, 6, 8, 9, and 11 displayed no evidence (0) of alignment to Sub-criterion C. Based on the data presented in Figure 25, there was little to no evidence of quality alignment to

inclusion of a schedule indicating how a PA was implemented across lessons and how the prior knowledge of students fit within a learning sequence.

Figure 25

Criterion 7 Sub-Criterion C Scaled Score Quality Ratings

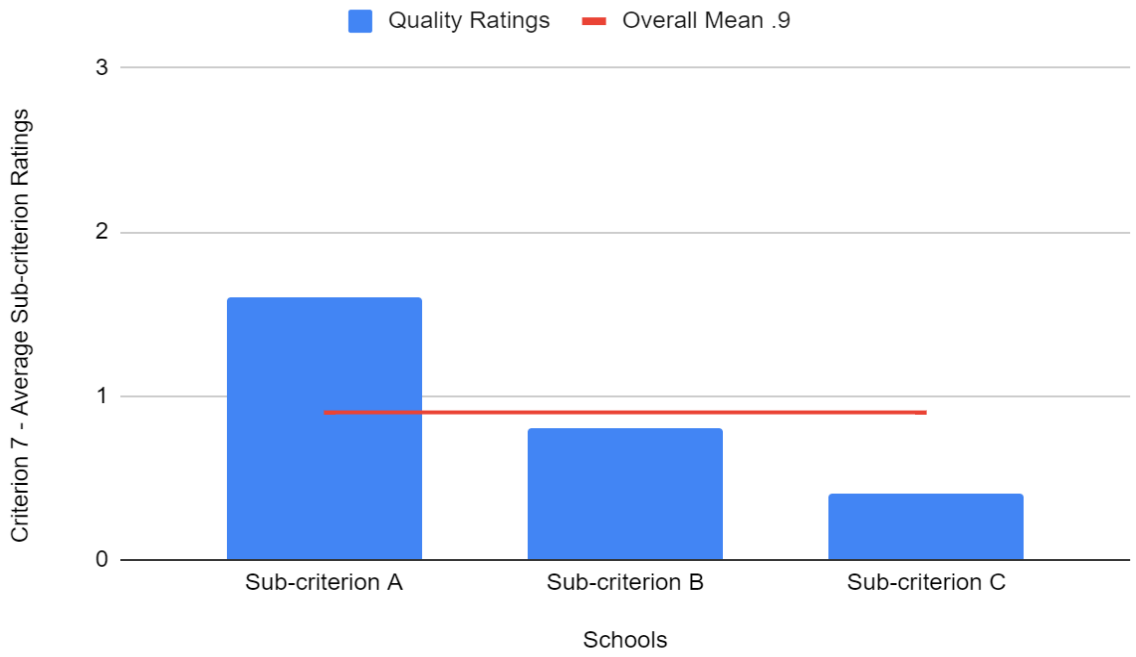


Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Figure 26 displayed Criterion 7's overall average ratings for Sub-criteria A, B and C. Accessibility to resources and materials needed for PAs and the inclusion of student-facing prompts, directions, and scoring tools (Sub-criterion A), showed data approaching partial evidence (1.6) of quality alignment. Data revealed that the complexity of PAs and whether the duration of implementation was realistic (Sub-criterion B) revealed less than limited evidence (.8). Lastly, the inclusion of a schedule indicating how the PAs were implemented across lessons and how the prior knowledge of students fit within the learning sequence (Sub-criterion C) revealed data barely above 0 (.4).

Figure 26

Criterion 7 Average Sub-Criterion Ratings



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Next, Table 16 revealed overall QCRT data by school and district. A narrative summary followed the table to highlight the main points of quality regarding PAs in PCPS.

Table 16

Quality Criteria Review Tool Summary Quality Data for Performance County Public Schools

School	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	Criterion 6	Criterion 7	Overall School PA Quality
	Standard/ ILO	Authenticity	Language Use	Success Criteria	Student Resources	Accessibility	Feasibility	
1	1.5	1	2.3	.3	.8	.5	.3	1
2	1.5	3	3	1	2.3	.3	.5	1.7
3	1.7	2	2.8	.3	1.7	1	.8	1.5
4	.5	1.5	1	.8	1.3	0	.5	.8
5	2.3	3	2	.5	2.2	.7	1.7	1.8
6	2.5	3	2.3	2	2.2	.8	.7	1.9
7	1.7	2.5	2.5	2.2	2.3	1.3	2.5	2.1
8	1.2	2	1.5	1	1.8	1	1.2	1.4
9	1.5	2.5	1.8	.7	1.3	1.3	.5	1.4
10	1	1.5	1.3	0	.8	1	.7	.9
11	2.8	3	3	1.2	2.7	2.3	.7	2.2
12	2.3	3	2.5	.3	1.8	1.3	1	1.7
M Rating	1.7	2.3	2.2	.9	1.8	1	.9	PCPS PA Quality Rating 1.5

Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence; PCPS = Performance County Public Schools; PA = Performance Assessment.

Overall Summary of Quality Ratings Criteria Data

Based upon the seven criteria that comprise the QCRT Quality Criteria Tool for Performance Assessments and aligned to the scaled score rubric, Criterion 2 (Authenticity) ranked highest in quality, as seen in Table 16 summary data. Evidence of authenticity is barely above partial evidence with a mean rating of 2.3. As seen in the Criterion 3 data, language use for expressing reasoning barely above partial evidence, as well, with a mean rating of 2.2 quality alignment and revealed the second highest rating in Table 16. When students express reasoning

they communicate feelings, thoughts, ideas and information. According to the QCRT, PAs should support students' language use by providing multiple means of accessing and using developmentally appropriate academic and disciplinary language, for example, text, video, audio, and oral. Criterion 5 is noted on Table 16 as the third overall highest quality rating with a mean rating of 1.8 and approached partial evidence. Criterion 1 is close behind and approached partial evidence with a mean rating of 1.7 quality alignment to standards and intended learning outcomes. Criterion 6, feasibility, revealed a mean rating of 1, showing limited evidence of alignment. Two criteria tied for the weakest rating; Criterion 4, success criteria, and Criterion 7, feasibility with mean scores of .9, revealing approaching limited evidence of quality alignment.

Next, Figure 27 is a bar graph indicating the QCRT summary quality data findings. The bar graph provided a visual representation of data from Table 16.

Figure 27

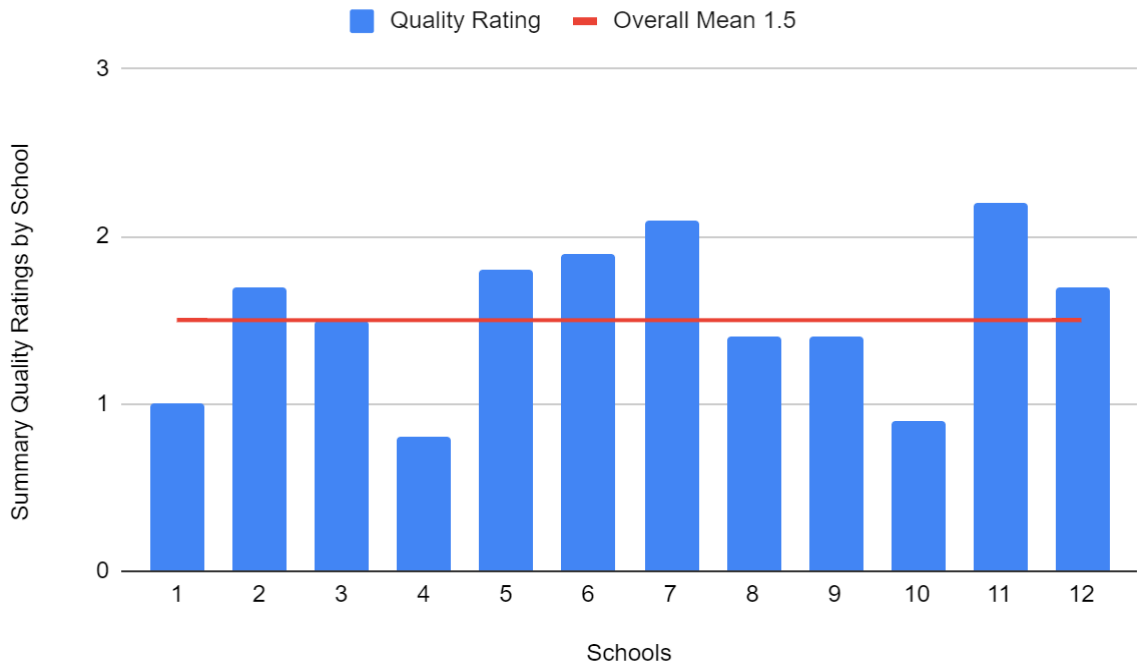
Summary Quality Ratings by Criteria

Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

Next, Figure 28 bar graph provided overall quality ratings by school.

Figure 28

Summary Quality Ratings by School



Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence

Displayed in Figure 28 is a bar graph indicating the summary quality data findings per school. The bar graph provided a visual representation of data from Table 16. Overall quality ratings by schools were as follows: School 11 had the overall highest quality rating of slightly above partial evidence with a 2.2; School 7 had the second highest quality rating of slightly above partial evidence with a 2.1; School 6 had an overall quality rating approaching partial evidence with a 1.9; close behind, school 5 had a quality rating approaching partial evidence with a 1.8; Schools 2 and 12 were slightly above limited evidence with overall quality scores of 1.7; School 3 approached partial evidence with a 1.5 overall quality rating; Schools 8 and 9 were

slightly above limited evidence with overall quality ratings of 1.4; School 1 had an overall quality rating of limited evidence with a 1; School 10 approached limited evidence with an overall score of .9; and school 4 approached limited evidence of overall quality with a score of .8.

The next section presented focus group data collected from 10 of the 12 middle schools in PCPS. School 3 and School 8 did not provide data. The knowledge, attitudes, thoughts, ideas, and opinions regarding the implementation of PAs by teachers in said school were valued in this process. Some focus group data may substantiate QCRT ratings.

Focus Group Data

Focus group data revealed how PAs were implemented in PCPS middle school social studies courses, as well as perspectives on the merit and worth of said PAs and recommendations for developing and using quality PAs moving forward.

Research Questions 2, 3, and 4, and answers to focus group questions led to findings that addressed the perceptions of USI and USII teachers

Findings for Research Question 2: How are performance-based assessments being implemented in USI and USII classes in PCPS?

In support of Research Question 2, participants were asked to answer focus group question one (FGQ.1) in which they were asked: What contributed to your preparation to design (or select) and implement PAs? Table 17 displayed categories and a breakdown of the frequency of responses among schools. Each check mark denotes a mention by one or more participants in each school.

Table 17

FGQ.1 Frequency of Responses among Performance County Public Schools

School	PBL	PLC	DBQ	Assorted Workshops	Graduate Courses	Summit Program
1		✓				✓
2	✓	✓	✓	✓		
4	✓	✓			✓	
5		✓	✓		✓	
6		✓	✓			
7	✓	✓	✓			
9	✓	✓				
10	✓	✓			✓	
11	✓	✓	✓	✓		✓
12		✓	✓			
Total	6/10 = 60%	10/10 =100%	6/10 = 60%	2/10 =20%	3/10 = 30%	2/10 = 20%

Note. PBL = Project-based Learning; PLC = Professional Learning Community; DBQ = Document Based Question

Commentary from a variety of schools was seen below and provided insight regarding how PAs were implemented in PCPS schools. The responses directly aligned to focus group question three (FGQ.3): *What contributed to your preparation to design (or select) and implement PAs?*

Project-Based Learning. Six schools specifically mentioned the 2014 PCPS countywide initiative when the expectation was that all K-12 schools be trained in performance-based learning (PBL). Initially, the training was provided by the Buck Institute and later by PCPS teachers and staff who were trained in Buck Institute's train-the-trainer workshops. While not all

participating schools mentioned PBL during focus group discussions, 8 out of the 10 schools in this study participated in PBL workshops. The exceptions were school 1 and school 4. However, one participant in school 4 did reference PBL training opportunities before coming to middle school:

> I've had some training in PBL when I was an elementary school teacher. I went to training at a local university for like 3 years in a row, so it was like a month of PBL training. And that was between the local university and PCPS.

Additional schools also referenced PBL. School 2 mentioned collaborating as a grade level, "When we had our PBL training in August 2014, we had time to meet as a grade level to plan our PAs."

School 7 said that PBL provided ideas about how to build cross-disciplinary content and skills. Additionally, Expeditionary Learning training was mentioned by School 9, and the fact they learned about PBL using the Expeditionary Learning method way before the Buck Institute came to PCPS. Finally, school 10 had one participant who referenced PBL training:

> I had 3 days' worth of PBL training. I don't think it really helped me. It was long and drawn out. It forced you to evaluate everything that you were doing, and maybe for new teachers it would be helpful, but for me it really wasn't. It was a matter of writing down what was already circulating. If that's helpful, then I don't see it. The training was too time-consuming. Pinterest helps a lot, and 28 years of teaching helps too.

Focus group question 1 asked participants to discuss what contributed to their preparation to design or select PAs. While only six schools made mention of countywide PBL training, it is known that a total of 8 schools out of the 10 participating in focus groups received the training.

Worthy of note is that the schools with the highest quality ratings participated in PBL training; schools with the lowest quality ratings did not.

Next, PLCs are described as a further contributor to the preparation and design of PAs.

Professional Learning Communities. All 10 middle school focus groups mentioned the professional learning communities (PLCs) as a factor that contributed to the preparation and implementation of PAs. School 12 mentioned the work of other teachers:

> I see what [other teachers] are doing in their classrooms and think, "I should be doing that with my kids." In our PLC we bounce ideas off of each other all of the time and I make changes and edits to my colleagues' ideas to best support my students.

School 9 referred to their school as a community of learners, "I really feel like what contributes for me is that community of sharing and learning new ideas and manipulating them to fit the content." Similarly, school 4 shared that building a rubric by collaborating works: "I kind of feel like my preparation has come more from collaborating together to come up with ideas and to come up with a rubric. I have not had much formal training."

School 1 provided evidence that demonstrated two very opposite scenarios in the preparation and implementation of PAs. One participant felt as though they already knew how to implement PAs without training, and another teacher in the same school had received targeted training in the implementation of PAs. These contrasting views are seen in the following two examples. First:

> I need to be prepared; I don't think I need to be taught how to do it. I don't need a class on it. It's just one of those things that you use what you know [students] need to know, and you create something using creativity. That's why we're in the profession. It's not

> one of those things that you can be taught to do. I just look at what needs to be done, needs to be taught, and make something based on that.

In the school 1 second example, the Summit Program, which is like a school within a school, typically meets in exclusive PLCs. Summit teachers have received extensive training on the implementation of PAs because PAs are at the core of the Summit program.

School 5 spoke to the value and fun of working in a PLC:

> There is value in a PLC because it gets exciting and actually fun sometimes if we get to talking and it gives us ideas. It's about bouncing ideas off of colleagues and working together. Somebody might be good with the computer side of it and somebody else with a different aspect of it. I think the final [PA] product benefits from having all that input.

PLCs were seen as a factor that contributed to the preparation, design and implementation of PAs. There was agreement amongst all schools participating in focus groups that collaborating in PLCs was helpful in gaining new ideas and perspectives about planning and implementing PAs.

Next, *The DBQ Project* is described as another factor that contributed to the preparation and design of PAs.

DBQ Project. Four schools mentioned attending multiple summer training sessions to learn about the implementation of document-based questions as contributing to their knowledge of preparing to design and implement PAs. The training was contracted with an outside organization known as *The DBQ Project*, as described in the literature review of this study. School 11 mentioned that DBQs provide a good foundation:

> The DBQ training really hit the mark. It really hits the curriculum and it just helps you. You can go from [ready-made PAs], or you can base your ideas from that, so you have a good foundation in how they created those.

School 12 mentioned DBQs as a support to their PLC:

> We have great resources like DBQs. Maliah attended a summer workshop from The DBQ Project. She brought those ideas to our PLC, and we brainstormed ways to best implement them. I like how we do not need to reinvent the wheel.

School 7 revealed benefits of attending workshops with their entire PLC, "We did DBQ training together and that helped a whole lot. We were able to take an idea or especially an historical document and flesh it out."

The DBQ Project workshops were described by four schools as foundational in recognizing and designing PAs. Participants spoke about the opportunities to use the DBQs as presented or modify to fit specific needs of students. Also mentioned was the value-added when attending DBQ workshops with PLC members. It should be noted that the two schools with the highest quality ratings participated in *The DBQ Project* workshops.

Next, Graduate school and assorted workshops are described as factors that contributed to the preparation and design of PAs.

Assorted Workshops and Graduate School. Three schools mentioned participation in workshops made available by the social studies curriculum specialist as contributing to the preparation to design and implement PAs. Some workshops were provided by organizations outside of PCPS, and other workshops were planned and implemented by the social studies curriculum specialist. Typically, the social studies curriculum specialist advertised workshops by sending emails to secondary department heads who then shared the opportunities with

department members. School 11 referenced several factors that contributed to their preparation to design and select PAs:

> I learned from The DBQ Project but also *Reading Like a Historian* through Stanford History Education Group (SHEG). I also stay in contact with places like the Holocaust Museum, the Virginia War Memorial, and Moton High School out in Farmville. I learned how to work with historical thinking standards like being able to do the interpretation, the analysis and comparing, and contrasting different viewpoints. I think a lot of that helps us prepare what we want to do for our performance-based assessments.

School 2 mentioned autonomy to choose:

> In the years I have known you as the curriculum specialist, I think it's good that you bring information to the table and you want us to take from it, but it's not the point like we are required to do X, Y and Z. So, I think it makes us feel more at ease to use trial and error, using best practices, and going forward to make [PAs] better.

School 2 also referred to the best timing to implement PAs:

> We're like the kids. We don't want things shoved down our throats. We want to feel like this is an opportunity. And you don't want to use PBAs for every little thing, because just like anything, they'll get sick and tired of it. But when you do use it, make sure it's the right time. We know our students, so that really allows us to maximize that opportunity.

School 5 mentioned graduate level courses:

> I will say that I have prepared through my master's program. The opportunity to take different classes, like ELL or SPED. That gives me different perspectives on how I can prepare and manage different assessments for different types of students. I am eager to implement those in my classroom.

School 10 also mentioned graduate level courses, “When I went to graduate school on education, they did an information session for us on PBL.”

Next, focus group question two (FGQ.2), along with Subquestions A through E, were presented in support of Research Question 2. FGQ.2 reads, How do you typically ensure a PA can produce appropriate and reliable outcomes? Subquestions were presented to focus groups to support FGQ.2. Subquestion A stated, *What is your typical practice of aligning a PA to the content that is taught?*

There were descriptive words and phrases that were gleaned from transcript statements provided by each of the 10 schools. The descriptive words and phrases were inductively analyzed and sorted into categories. Table 18 reveals a breakdown of the frequency among schools regarding responses. Each check mark denotes a mention by one or more participants in each school.

Table 18

FGQ.2 Frequency of Responses among Schools for Subquestion A

School	Backward Design	Align to SOL
FG.S1	✓	✓
FG.S2		✓
FG.S4		✓
FG.S5	✓	
FG.S6	✓	✓
FG.S7	✓	
FG.S9		✓
FG.S10	✓	✓
FG.S11		✓
FG.S12	✓	✓
Total	6/10 = 60%	8/10 = 80%

There was commonality among schools' responses with respect to how PAs are aligned to the content. Commentary is provided below from some of the schools regarding Subquestion A.

School 4 described alignment to the skills and standards, "One of the things I do is look at the standard that we're teaching and look for the verb of what the kids have to be able to do, like describe or identify, and then go from there." Similarly, school 10 mentioned:

> I honestly just look at the standard. I'll see what the standards are asking, and then if I am creating it, or if I'm implementing someone else's, I will actually put the standard in my own words, and then have the students see that, and say, this is why we are doing this part; [PA] aligns with "X" standard. So, I want them to see that, and then I want to see

> that, because it allows me to ensure that it is being aligned correctly. And it does force me to analyze the language.

Additionally, school 12 described backward design:

> "What is my end result?" is my go-to question. Being fresh out of grad school, we learned backward design. So, what do I want them to take away that's most important, and how do I accomplish that?

Similarly, school 6 actually created the final product:

> We [PLC] worked together to actually make the final product for the unit to determine if we thought our students would be able to make it. Later, we used our example to show the students.

Next, focus group responses are provided to Subquestion B to support FGQ.2.

Subquestion B stated, *Do you determine if essential resources for a PA are available for all students before implementing a PA? If so, how?* There were descriptive words and phrases that were gleaned from transcript statements provided by each of the 10 schools that were inductively analyzed and sorted into categories. Table 19 reveals a breakdown of the frequency among schools regarding responses. Each check mark denotes a mention by one or more participants in each school.

Table 19

FGQ.2 Frequency of Responses among Performance County Public Schools for Subquestion B

School	Readily Available Resources
FG.S1	✓
FG.S2	✓
FG.S4	✓
FG.S5	✓
FG.S6	✓
FG.S7	✓
FG.S9	✓
FG.S10	✓
FG.S11	✓
FG.S12	✓
Total	10/10 = 100%

There was consensus among schools' responses with respect to providing essential resources for all students. All 10 schools provided varying ways for students to access readily available resources. Commentary was provided below from some of the schools regarding Subquestion B.

School 4 told students to ask teachers for resources they did not have: "When we did the artifact project we made sure that students knew that if they didn't have things at home, that they could ask [teachers]. We gave them a date to ask us for the resources that they needed."

School 5 provided the ways they helped students with disabilities:

> Making sure that you have the resources for students you have to accommodate is not something everyone thinks about. If you have a video, it can be resourceful for English

> Language Learners if there are subtitles. I had a student with dyslexia and the video worked well for all students but this one. So, I got dyslexic friendly font subtitles. These types of resources available within the school can make a big difference for students.

School 6 does not let students take materials home:

> We provide everything for them and everything is done in the building. Parents have asked me if they can take things home and I just say no because I know I would never get the project back, plus, I do not want parents doing [students'] work.

School 10 supplies the exact websites: "I provide the resources. And if it requires online research, I will actually provide the students with a list on a Google doc of the approved websites they can go to answer questions."

School 9 works with students who do not have Internet access:

> I give out poster boards and supplies if they need something. We usually give them [students] some choices of things, so if they don't have Wi-Fi at home, we give them enough choices in the projects so that they can do one that doesn't need it.

Next, focus group responses were provided to Subquestion C to support FGQ.2. Subquestion C stated, *How do you determine that all students clearly understand the communicated procedures of a PA?* There were descriptive words and phrases that appeared that were gleaned from transcript statements provided by each of the 10 schools. The descriptive words and phrases were inductively analyzed and sorted into categories. Table 20 revealed a breakdown of the frequency among schools regarding responses. Each check mark denoted a mention by one or more participants in each school.

Table 20

FGQ.2 Frequency of Responses among Performance County Public Schools for Subquestion C

School	Provide Examples	Rubric/ Checklist	Written Directions	Teacher Feedback	Repeating/ Step-by-Step Directions
FG.S1		✓		✓	✓
FG.S2			✓		✓
FG.S4	✓				✓
FG.S5	✓			✓	
FG.S6		✓	✓		
FG.S7	✓	✓	✓	✓	
FG.S9	✓	✓	✓		✓
FG.S10	✓	✓	✓		✓
FG.S11		✓	✓	✓	
FG.S12	✓	✓			
Total	6/10 = 60%	7/10 = 70%	6/10 = 50%	4/10 = 40%	5/10 = 50%

Seven out of 10 schools mentioned making sure that all students understood the communicated procedures by providing a rubric/checklist, while schools provided examples and written directions for students. Five schools repeated step-by-step directions, and three provided teacher feedback. All schools mentioned implementing two of more of the strategies seen in Table 20. Commentary was provided below from some schools related to Subquestion C.

- School 2 provided directions in a variety of ways: “Some classes need more repetition than others. We’ll say it, we’ll have it posted in the room, we’ll have it online whether it’s Canvas or Google Classroom. So, it’s digital, it’s verbal and it’s visual.”

- School 6 mentioned typed out directions: “Each student has a packet with all of the directions typed out. They have checkboxes where they start here, okay you’ve done this—check—move to the next thing on the list. I don’t think I’ve had an issue with them not knowing really what to do next.”
- School 9 also mentioned written directions:

  > We give students written directions. We try to make sure we get it posted in various places. Directions get vetted over the years to some degree. We have to make sure we tweak the directions for students from year to year so that they are clear.

- School 10 added stopping at each step to review directions:

  > Not only do I give the rubric and read over the rubric, I actually stop at every step to ensure that the students understand directions. A lot of times I will provide examples to make sure they understand. I give time for students to ask questions as well.

Next, focus group responses were provided to Subquestion D to support FGQ.2. Subquestion D stated, *How do you ensure that students have a range of opportunities for success on a PA?* There were five descriptive words and phrases that appeared on Table 21 beside FGQ.2, Subquestion D that were gleaned from transcript statements provided by each of the 10 schools. The five descriptive words and phrases were inductively analyzed and sorted into categories. Table 21 revealed a breakdown of the frequency among schools regarding responses. Each check mark denoted a mention by one or more participants in each school.

Table 21

FGQ.2 Frequency of Responses among Performance County Public Schools for Subquestion D

School	Student Choice	Teacher Feedback	Differentiated Scoring	Student Reflection and Re-do	Clear Expectations
FG.S1	✓	✓		✓	
FG.S2	✓	✓	✓		
FG.S4	✓				
FG.S5	✓			✓	
FG.S6	✓		✓		
FG.S7		✓			
FG.S9	✓			✓	
FG.S10	✓				
FG.S11		✓			✓
FG.S12		✓	✓		
Total	7/10 = 70%	5/10 = 50%	3/10 = 30%	3/10 = 30%	1/10 = 10%

Student choice was the most common response regarding the best way for students to achieve success, mentioned by 7 out of 10 schools. Teacher feedback was mentioned by five schools, while three said that student reflection and allowing students to re-do their work helped students achieve success. Three schools mentioned the use of differentiated scoring, and one school referred to providing clear expectations.

Commentary was provided here from some of the schools regarding Subquestion D. School 5 mentioned student choice:

> I think student choice, in what the end product is, to a degree is important. So is the choice of who they can work with and what their topic is. In order to get them interested

> and into something, they have to have some say in what they're doing so it's just not coming from [teachers].

School 11 discussed teacher feedback and re-dos, "If students don't do well the first time, we actually sit down and go over what is missing. We give them feedback and allow them to try again until they get it." School 12 also mentioned feedback:

> Our USI PLC has focused on providing direct feedback to each child. Instead of waiting until the end when students turn in their project, we provide feedback throughout the process. We also provide check-in dates for students to have portions of the work completed.

School 2 provided an example of balanced assessments:

> Typically, when we have PBAs, it's not like the only thing that we are working on. A lot of times we'll have mini-lessons or a test as we go through it so that it provides equity for everything to learn and ask questions in different settings, and have the information in front of them. We feel this is a way we can ensure success on any given PBA.

Next, focus group responses were provided to Subquestion E in support of FGQ.4. Subquestion E stated, *How do you use the results of the PA to directly impact the teaching and learning of your students?* There are five descriptive words and phrases that appeared on Table 22 beside FGQ.2, Subquestion E that were gleaned from transcript statements provided by each of the 10 schools. The five descriptive words and phrases were inductively analyzed and sorted into categories. Table 22 revealed a breakdown of the frequency among schools regarding responses. Each check mark denoted a mention by one or more participants in each school.

Table 22

FGQ.2 Frequency of Responses among Performance County Public Schools for Subquestion E

School	Find Weaknesses & Reinforce Learning	Reflections & Modifications
FG.S1	✓	
FG.S2		✓
FG.S4	✓	
FG.S5		✓
FG.S6	✓	
FG.S7	✓	
FG.S9	✓	
FG.S10	✓	
FG.S11	✓	
FG.S12		✓
Total	7/10 = 70%	3/10 = 30%

To influence the teaching and learning of students based upon the results of PAs, 7 out of 10 schools reviewed the student weaknesses to reinforce learning. Three schools reflected on the outcomes and made modifications to the PA.

Commentary is provided here from some of the schools regarding Subquestion E School 10 reflected on fine-tuning skills:

> Performance assessments help you get really finely tuned to the strengths and weaknesses of students. Whether it's their reading or writing level, or their comprehension level, you can take that and apply it in the future when you are working with certain groups of students, to meet the needs that they have, or, to push them in a higher direction than where they currently are.

School 9 focused on spiraling back:

> Where I find weaknesses, I like to spiral back. I look for trends within data that's been collected and say, hey, I must have taught this terribly wrong because nobody got that. Or, we all seem really good at this lesson and to apply it to a future lesson, because it really was helpful on that piece, more so than other strategies.

School 4 mentioned how results drive instruction:

> The results of a PA drive instruction. If [a student's] product doesn't show that they learned the standard, you know you need to revisit it. Or, if it shows that's something you could go even more in depth with if they're excited about it. I feel like assessment drives our instruction.

Finally, school 1 offered that self-reflection is important, "I always self-reflect too about what worked and what didn't work. I try to fix whatever needs to be, like in the rubric or directions. The reflections influence my future teaching." The next section will focus on research question 3.

Findings for Research Question 3: What are the USI and USII teachers' perspectives on the merit and worth of the performance assessment process in PCPS?

In support of Research Question 3, participants were asked to answer focus group questions 3 and 4 (FGQ.3 and FGQ.4). FGQ.3 is, *Do you support the move toward the use of performance assessments to measure student learning? Why do you feel this way?* Nine descriptive words and phrases appeared on Table 23 that supported the move toward the use of PAs to measure students' learning and three points that showed minimal support. The descriptive words and phrases were inductively analyzed and sorted into categories. Table 23 categorized the nine points into three columns to reveal the breakdown of the frequency among schools

regarding responses to FGQ.3. Each check mark denoted a mention by one or more participants in each school.

Table 23

FGQ.3 Frequency of Responses Among Schools in Support of Performance Assessments

School	Show what students know by way of deeper thinking skills	Provides balanced and equitable assessment types and opportunities for differentiation & to measure student growth	Motivates student achievement through freedom of expression and less stress to pass a standardized test
FG.S1	✓	✓	
FG.S2	✓		
FG.S4	✓	✓	✓
FG.S5	✓	✓	✓
FG.S6	✓	✓	
FG.S7	✓	✓	
FG.S9	✓	✓	
FG.S10			✓
FG.S11	✓	✓	
FG.S12		✓	✓
Total	8/10 = 80%	6/10 = 60%	4/10 = 40%

Table 23 showed the frequency of responses in support of PAs to FGQ.3. The most common answer, shared by 8 out of 10 schools, focused on how PAs allowed students to show what they knew by engaging deeper thinking skills. Six schools mentioned that PAs provide an equitable balance of assessment types and the opportunity to differentiate. Four schools claimed that PAs motivated student achievement through freedom of expression and less stress to pass a standardized test.

Commentary was provided here from some of the schools regarding FGQ.3. FG.S4 values the move toward equity:

> I think that it's more of a move toward equity. We have students who all come from different backgrounds, and all learn in different ways, and this is just an opportunity to give them different ways to show what they know.

FG.S5 speaks to lessening pressure on teachers and students:

> I definitely support the move towards PAs because it alleviates the pressure of the teacher as well as the students to do well to pass a [SOL] test. I also think they are better because not every student is the same...you're able to differentiate your PAs based upon each student's strengths and weaknesses.

FG.S10 mentioned PAs as a motivator:

> I did my student teaching in New York City, and one of the requirements for every school in New York was to have at least one performance-based assessment per quarter. So, I got to see that in practice, and got to see the benefit of it. To me, it serves as an incredible motivator for student achievement.

FG.S11 recognized deeper thinking skills:

> We are noticing that students are weak with writing skills but I still love that it makes them go deeper. It actually makes them have to think through things. They're not just memorizing one thing; they're looking at two sides of an issue and they get excited about it. One of my favorite things was having the debate, and the debate was great, but then the next day we switched sides. They had to figure out what the other side would think. I let them choose the side they wanted, they prepared, and then they had the next day to do the opposite. That really shook some up, but it really made them go deeper. Even as a

teacher, it makes it more enjoyable. Kids learn in all different ways. Some kids are awesome at test-taking and some kids are terrible test-takers. Through PAs you can really see who actually understands and can synthesize and not just regurgitate, and actually come up with their thoughts about an exciting topic. It's just a neat way to teach.

Next, Table 24 categorizes the frequency of responses among schools with minimal support of PAs. Each check mark denotes a mention by one or more participants in each school and suggests very little opposition to PAs.

Table 24

FGQ.3 Frequency of Responses among schools with Minimal Support of Performance Assessments

School	Take too much time to implement with large numbers of students	Students are not self-aware and don't take PAs seriously
FG.S1	✓	
FG.S2		
FG.S4		
FG.S5		
FG.S6		
FG.S7		
FG.S9	✓	✓
FG.S10		✓
FG.S11		
FG.S12	✓	
Total	3/10 = 30%	2/10 = 20%

Based upon responses from 3 out of the 10 schools, there were two main concerns that appeared. One focused on the fact that PAs take too much time to complete and score with large numbers of students. Second, two schools felt that students were not self-aware enough and did not take PAs seriously.

Next, a second question in support of Research Question 3 was FGQ.4: *Do you value the opportunity to choose or create performance-based assessments to measure student learning?* Thirteen descriptive words and phrases on Table 25 that responded to said question. The thirteen descriptive words and phrases were inductively analyzed and sorted into categories. Table 25 categorized the points and frequency. Each check mark denoted a mention by one or more participants in each school.

Table 25

FGQ.4 Frequency of Responses Among Schools Valuing the Opportunity to Choose

School	Choice motivates teachers to take ownership and honors them as professionals	Can discern strength and weaknesses of students so that weaknesses become strengths with teacher feedback	Promotes collaboration among teachers	Tailor to the needs of students in order to differentiate
FG.S1	✓			✓
FG.S2	✓	✓		✓
FG.S4	✓			✓
FG.S5	✓	✓		✓
FG.S6		✓		✓
FG.S7		✓	✓	
FG.S9	✓	✓	✓	
FG.S10	✓	✓		✓
FG.S11	✓		✓	✓
FG.S12		✓		✓
Total	7/10 = 70%	7/10 = 70%	3/10 = 30%	8/10 = 80%

Answers to FGQ.4 revealed that participants in this study valued the opportunity to choose or create PAs. Responses from 8 of the 10 schools revealed that having choice allows them to tailor the needs of their students in order to differentiate. Choice also motivates teachers to take ownership of their work along with honoring them as professionals, as revealed by 7 of the 10 schools. Seven schools felt that because they can discern the strengths and weaknesses of their students, they can design PAs to best fit their needs, to include teacher feedback. Three schools value the opportunity to choose because it promotes collaboration among teachers.

Commentary was provided from some of the schools regarding FGQ.4. First, School 1 suggested, “It would be nice to have a combination of PAs I create and a bank to choose from. Since we have a choice we can tweak the ones from the bank to make it our own.” School 4 mentioned teachers as professionals: “I value the opportunity to choose because I think it lends a hand to showing that the teachers are professionals; we’re able to connect with the curriculum and choose what is best for our students.” School 7 mentioned team time to plan “Every class every year is different. So often we’ll sit down during team time and might say, “there is no way this [PA] will work.” Then, we work together to alter it.”

School 9 referred to autonomy:

> I like the ability and autonomy to recognize strengths and weaknesses for my students from formative assessments that allow me to turn around and say, based on what I’ve noticed, maybe if we work on it this way, I’ll see those weaknesses become strengths.

School 5 values the opportunity to differentiate for certain groups of students:

> Because student populations are not the same everywhere throughout the county teachers are able to devise something that is going to be more relatable or engaging to certain groups of students, or even the different classes among your six classes. Something might work better with one group than another and having choice gives you more autonomy to do something that would be a more valuable learning experience.

Findings for Research Question 4: What are USI and USII teachers’ recommendations for developing and using quality performance assessments in PCPS moving forward?

In support of Research Question 4, participants were asked to answer focus group Question 5 (FGQ.5), *Do you feel that the performance-based assessments you implement with your students reflect what students know and are able to do with what they know? If so, how? If*

you are not sure, what ideas do you have that could assist in planning a professional development opportunity to best support the needs of other teachers who may not be sure?

Seven summative points that appear in Table 26 in favor of teacher-made PAs reflected what students know and were able to do with what they know gleaned from transcript statements provided by each of the 10 schools; six summative points appear in contrast. Summative points were categorized and labeled and appear in Table 26, which shows a breakdown of response frequency across schools. Each check mark denoted a mention by one or more participants in each school. Also seen in Table 26 are seven suggestions provided by teachers in response to FGQ.5a: *What ideas do you have that could assist in planning a PD to best support your needs and the needs of other teachers in regards to implementing PA?* These responses help to better support the implementation of PAs in PCPS.

Table 26

FGQ.5 Frequency of responses among schools regarding varying support of PAs reflecting what students know and are able to do with what they know

	Support		Minimal Support	
School	Genuine Student Interest	Deeper Learning Skills	Depends upon the Students	Depends upon the Rubric
FG.S1		✓		
FG.S2	✓		✓	
FG.S4			✓	
FG.S5				
FG.S6				
FG.S7				✓
FG.S9	✓		✓	
FG.S10	✓			✓
FG.S11				
FG.S12				
Total	3/10 = 30%	1/10 = 10%	3/10 = 30%	2/10 - 20%

A nominal number of responses were provided by only 6 of 10 schools and are seen in Table 26. PAs that provide students an opportunity to engage in topics they are truly interested in show what students know and are able to do with what they know, as revealed by three schools. One school mentioned the deeper learning skills that engage students in a PA show what students know and are able to do with what they know. In contrast, three schools said that it really depends upon the effort of the students to provide a true sense of what they know and are able to do with what they know. In addition, two schools refer to the importance of clear expectations on the rubric in order to see what students know and can do with what they know. FGQ.5 was the

most challenging question for schools to interpret, as evidenced by the lack of available pertinent responses.

Commentary was provided from some of the schools regarding FGQ.5. First, School 2 mentioned building the interest of the students:

> It depends on your students and how much interest and familiarity they have with the topic. I really had to build my students emotional connections to topics like World War II so that they are ready to dive into a PA. If I were to be the head of social studies in America, I would instruct all teachers to build it up, not just throw a full-blown PA in front of students every year.

School 9 said, "I think PAs don't just reflect what students know, but they learn while they're doing it." School 4 focused on the effort of the students:

> In my opinion, and it's unfortunately kind of negative, not everyone does the project and a lot of students don't really put forth effort. So, I have often found it difficult to really see if they've learned what I've intended for them to learn. And I have also found that I end up lowering my standards so that a lot of students do not fail.

School 10 stated that PAs are not for all students:

> I think in some cases PAs reflect what students know and can do, however, I still think that PAs aren't for all students. I think for the majority of students it does work, but there are some students that are still developing into that level where they have that higher level of thinking to be able to apply their knowledge and actually create something from what they've learned, instead of remembering or reciting it.

In support of Research Question 4 and FGQ.5, is a sub question, *What ideas do you have that could assist in planning a PD to best support your needs and the needs of other teachers in*

regards to implementing PAs? Five summative points suggested ways to support teachers with the implementation of PAs:

- provide a bank of PAs aligned to content topics and skills
- offer a countywide gallery walk showcasing teacher PAs
- provide more concrete expectations about the implementation of PAs
- arrange opportunities to learn about how to create both big and small PAs
- offer culturally responsive professional development in the planning and implementation of PAs

Some schools were forthcoming with their ideas and suggestions about how to support the implementation of PAs. The five suggestions appeared to be relevant and reasonable in their scope. The next section briefly reflected upon data collection outcomes and the inductive development of relevant themes.

Summary

In preceding sections of Chapter 4, quantitative and qualitative data, as well as summaries, were explained in detail. Quantitative data analysis findings, regarding the degree of quality seen in teacher-selected performance assessments, revealed an overall mean score of 1.5. A score of 1.5 indicates the quality of said assessments lies between limited to partial evidence based on the Summary Quality Ratings by Criteria Scale. Qualitative data analysis findings resulted in 11 themes as follows:

1. Teachers demonstrated widely varying degrees of understanding creating quality PAs.
2. Teachers acknowledged the practice of aligning PAs to the big ideas and skills provided in the Virginia SOL.
3. Teachers made essential resources available for students when assigning a PA.

4. Teachers had varied ways of communicating PA procedures with students.
5. Teachers valued student choice.
6. Teachers used the results of PAs to reinforce learning.
7. Some teachers valued PAs because of the differentiated opportunities for students to show what they know.
8. Some teachers did not value PAs because they take too much time and students are not developmentally ready.
9. Teachers valued the opportunity to choose or create their own PAs.
10. Some teachers believed that PAs provided an accurate representation of what students know and are able to do.
11. Based on teacher feedback, a bank of PAs, as well as varied PA learning opportunities, will enhance the implementation of PAs in PCPS.

While the prior themes did draw mostly positive conclusions regarding the use of PAs, the quality data of said PAs did not distinctly align to these perceptions. This topic was further addressed in Chapter 5 along with additional conclusions relative to the purpose of the study. Further, responsiveness to key issues about the quality of teacher-selected performance assessments were addressed. Finally, coalescence of criterial measurement and interpretation led to an understanding about quality and implications for practice and future research.

CHAPTER 5

SUMMARY, DISCUSSION, AND IMPLICATIONS

A brief summary of the study's purpose with relevant conclusions reached as a result of the data analysis is presented. Then, findings are investigated and discussed, resulting in opinions developed in relation to those findings. Finally, implications are explored regarding teachers' professional development along with suggestions for future research.

Summary of the Study

Stake (2004) posited that responsive evaluation is "a search for and documentation of program quality and uses both criterial measurement and interpretation" (p. 89). Founded on Stake's claim, notable conclusions were drawn based upon quantitative and qualitative data derived from findings in this study. Hence, the purpose of this responsive evaluation was to search for and document quality features of teacher-selected performance assessments (PAs). The instrument used to determine said quality was the Virginia Quality Criteria Tool for Performance Assessment (QCRT). Sample PAs were selected from two middle social studies courses: United States History to 1865 (USI) and United States History 1865 to the Present (USII). At the outset, QCRT quantitative data were collected concurrently with qualitative focus group data and analyzed independently. The quantitative data were derived from the QCRT ratings of 24 PAs from 12 middle schools. The qualitative focus group data were obtained from recordings and transcripts representing 10 of the 12 middle schools. Qualitative data analysis findings resulted in 11 themes that were compared side-by-side to quantitative quality ratings of teacher-selected PAs in an effort to better understand teachers' perceptions of quality. Focus

group responses were used as the means of addressing Research Questions 2, 3, and 4. The next section provides a discussion and in-depth reflection regarding the quality of teacher-selected performance assessments and next steps to improve the quality of said assessments.

Discussion

The findings of this study on the sources of quality of teacher-selected performance assessments matter because a lack of quality, teacher-selected performance assessments could mean that some students may not get an opportunity to move beyond a narrow set of curriculum and skills. Moving beyond a narrow set of curriculum and skills was one thing the Virginia General Assembly had in mind when legislative mandates amended the *Code of Virginia* to eliminate several SOL tests (VDOE, 2014). Additionally, quality performance assessments have the potential to inform instruction in order to ensure that students are making adequate academic progress. While this study may have been limited in size and scope to address issues of quality performance assessments as a problem of practice, the consistency reported among schools in regards to the weakest and strongest quality criteria was noteworthy.

After the VDOE replaced certain SOL assessments with alternative assessments in 2014 and provided the autonomy for school divisions in Virginia to develop their own, the idea of designing a study to review the quality of said assessments seemed to be a logical next step. School division leaders in PCPS chose to design and implement a portfolio approach to assessment with guidelines that permitted teachers and students to decide what alternative assessments would be showcased in the portfolio. By 2016, however, the VDOE expected school divisions to use at least one performance assessment in classrooms where a standardized test was replaced with an alternative assessment (VDOE, 2016b). Introduced in 2017 and updated in 2019, *The Virginia Quality Criteria Review Tool for Performance Assessments (QCRT)* was

presented to school divisions as a detailed guide to help school divisions recognize, create, or revise performance assessments that measured the application of content knowledge and skills. The criteria were designed to support comparability in rigor and quality across the state to achieve consistent expectations for Virginia students (VDOE, 2019b). It was suggested that the use of this tool, as well as the common rubrics that were also introduced, would bring into existence a more objective and equitable means of assessing students. The most current VDOE Superintendent's Memo #214-21, updated on August 6, 2021, provides a definition of authentic performance assessments as follows:

> Performance assessments generally require students to perform a task or create a product that is typically scored using a rubric. Authentic performance assessments often include tasks that mirror those that might occur in a "real-life" situation and/or are authentic to the academic discipline (Attachment A, para 6).

The VDOE expected school divisions to continue using the QCRT "during the development of new performance assessments and/or the revision of existing tasks in order to ensure that all students have access to quality assessments aligned to the SOL (Attachment A, para 7)." This criteria tool communicated the quality features that should be present in a PA to provide consistent achievement expectations for Virginia students.

The QCRT was presumed an adequate instrument to characterize the quality of performance assessments for use in this study. This presumption was based on the fact that the set of expectations was established by the VDOE, and incorporated elements of work being done at the Stanford Center for Assessment, Learning and Equity to develop assessment systems that support more meaningful forms of student learning. As the researcher in this study, I recognized the validity of the QCRT in a variety of ways; I saw opportunities for alignment of PAs to

standards and intended learning outcomes (criterion 1); there were varying degrees of authenticity seen in PAs (criterion 2); I could pinpoint the incorporation of language within PAs for the purpose of expressing reasoning (criterion 3); PAs had a variety of success criteria (criterion 4), as well as, varying ways that PAs presented student directions, prompts and access to resources/materials (criterion 5). I could also make judgments as to the accessibility (criterion 6) and feasibility (criterion 7) of PAs based upon the first five criteria. In other words, items on the QCRT were represented in the PAs.

While the validity of the QCRT was evident, I noticed a challenge related to the use and reliability of the tool. Establishing interrater reliability during the quality reviews of each of the 24 PAs was imperative due to a mixed message regarding the implementation of the tool. On the first page of the tool there was a paragraph that communicated two different ways to interpret the use of the tool, as follows:

> ***It is not expected that every quality performance assessment will have full evidence that corresponds to each description within the Virginia Quality Criteria Review Tool.*** For example, it is not expected that a task include each of the deeper learning competencies listed in criterion 1C (i.e., mastering rigorous academic content; learning how to think critically and solve problems; working collaboratively; communicating effectively; directing one's own learning; and developing an academic mindset), but a quality task should provide the opportunity for students to demonstrate one or more of these competencies. Additionally, the criteria may be considered in any order that suits the division's needs and purpose. School divisions are encouraged to use this tool not only as they develop new performance assessments, but also as they evaluate the quality of existing tasks. ***If a task does not fully meet the expectations detailed for that criterion***,

> the school division should use the evidence or rationale to modify and improve the task before using that task with students (p. 1).

The first sentence in the above paragraph affirms the expectation that *not every quality performance assessment will have full evidence* of each criterion in the tool. However, the last sentence states that if the task does not *fully* meet expectations of the criterion, it must be modified to do so. Due to the inconsistent wording, the process of establishing interrater reliability for each PA quality review became a critical step.

Presented next, is an illustration to exemplify why a study on the potential lack of quality, teacher-selected performance assessments matters.

Sample Assessment

One example of a weak teacher-selected PA reviewed for quality in this study, narrowly reflected alignment to deeper learning and critical thinking. In this sample assessment, students were asked to design a travel brochure encouraging tourists to visit regions of the United States. Students were told to list states in the region and include images of products, landmarks, state flowers, birds, and flags. While the project did have a few quality features, such as alignment to state content standards and enabling students to present what they knew about the regions in a way different from a traditional multiple-choice assessment, the performance assessment did not go much beyond simple recall. The scoring tool was a checklist, requiring all students to create the same product with required information; there was no language to describe learning intentions or how a student response should demonstrate what they learned. There was no place to provide feedback to students about their work and how it might be improved. While the brochure assignment was most assuredly an alternative to a multiple-choice assessment, deeper learning competencies such as thinking critically and solving problems were not realized.

The brochure assessment was a classic example of a teacher-selected performance assessment lacking in capability to move beyond a narrow set of curriculum and skills to achieve an in-depth understanding of the overall curriculum and help inform instruction. A suggested way to improve one aspect of deeper learning in said brochure assessment might be to begin by providing students with a scenario where a group of friends are planning a vacation somewhere in the United States but cannot agree upon where to go. This scenario provides purpose and a real-world opportunity for students to collaborate with a group of friends and to solve a problem. As well, the deeper learning competency of collaboration is one of the five competencies mentioned in the "Profile of a Virginia Graduate," which describes the knowledge, skills, attributes and experience identified by employers, higher education and the state Board of Education as critical for future success (VDOE, 2016b). As seen in this example, a small modification to the brochure assessment yields a thoughtful opportunity to go beyond simple recall and apply knowledge in the design of an authentic product.

The next section will reinforce why teacher professional development is necessary to help improve upon the lack of quality performance assessments.

Professional Development

Another reason a study on the potential lack of quality, teacher-selected performance assessments matters is because identifying the weaknesses, as seen in the aforementioned example, can help to better align, plan, and implement professional development for teachers in order to improve the quality of said assessments. For more than 20 years, teachers have prepared students to pass standardized multiple-choice tests. Quality performance assessments were most likely not in demand during those years, nor professional development workshops intended to help teachers design them.

As I searched for empirical studies to support my dissertation topic, I noticed a scarcity of research targeting performance-based assessments between the years of 1998 to 2005. This lack of research may suggest that teachers today lack the skills to recognize, create and score quality performance assessments (Bland & Gareis, 2018). Frey and Schmitt (2010) noted that teachers often rely on assessments created mainly by others, such as textbook companies, and therefore may benefit from training on how to recognize quality features of performance assessments. Curry and Smith (2017) observed in their 3-year longitudinal study between 2012 and 2015 that assessment practices in social studies classrooms were more likely to include selected-response items rather than performance-based techniques.

In PCPS, I made similar observations. During visits to USI and USII classrooms after SOL tests were removed and replaced with the alternative portfolio assessments, I noticed that many teachers continued to instruct and assess students as if there would be a standardized test for students to pass in the spring. In general, while most teachers were enthusiastic about moving away from the SOL test, many did not change their style of instruction and assessment to include performance-based tasks and assessments for deeper learning. These same schools became participants in this study and during focus groups communicated support for the use of performance assessments as a way to measure student learning for three reasons: they show what students know; they provide balanced and equitable assessment types and opportunities for differentiation and measurement of student growth; and they motivate student achievement through freedom of expression and less stress to pass a standardized test. It appeared that while teachers supported the move to performance-based assessments, the expertise necessary to design and implement performance assessments, that went beyond a narrow set of curriculum and skills designed to inform instruction, was generally very weak. In other words, teachers may

have assumed they were implementing quality performance assessments when in essence, they were not. While teachers could speak to best practices in the implementation of PAs, the application did not always transfer to the actual PAs implemented with students. Teachers were aware of performance assessment characteristics and practices and were actually implementing varying degrees of best practices in an initial stage of what is called *consciousness raising* (Moss & Brookhart, 2008).

Next, PCPS district-wide, project-based learning professional development is addressed as a way that some teachers learned how to design and implement PBAs.

Project-Based Learning

An event that occurred in PCPS, simultaneous to SOL tests being replaced, was a district-wide initiative to educate all K-12 teachers in the practices of project-based learning. The initiative boosted the knowledge of teachers to design and implement quality performance assessments. The timing of the initiative was especially fortuitous for teachers of courses where an SOL test was replaced with an alternative assessment. While there was an expectation that all schools participate in this training, not all schools did. Schools 1 and 4 in this study did not participate in district level performance-based learning (PBL) training and showed markedly weaker quality ratings on teacher-selected performance assessments than schools who participated in the district-level PBL training. The overall weakest quality performance assessment ratings were schools 1 (1), 4 (.8), and 10 (.9). While school 10 did participate in the PBL training initiative, focus group feedback generally concluded that the opportunity was of little value. It is important to keep in mind however, while Schools 1, 4, and 10 had the lowest quality performance assessment ratings, none of the schools participating in this study had ideal performance quality assessment ratings. Therefore, the data derived from this study will be used

to intentionally plan professional development for all schools to improve the practice of recognizing, creating and implementing quality performance assessments.

The following sections discussed the quality ratings seen in this study in more depth. First, criteria with the lowest quality ratings in this study will be reviewed to help prioritize what is lacking in teacher-selected PAs to inform a plan for improvement. Second, criteria with the highest quality ratings will be considered for the purpose of a more in-depth exploration of how or why these criteria might have had higher quality ratings as another way to support a plan for improving PAs. Lastly, the remaining criteria quality ratings will be reviewed with a similar purpose in mind.

Criterion 4: Success Criteria

Based on the data noted in chapter 4 and seen again in Table 16, Criterion 4 tied for the lowest QCRT rating, below limited evidence (.9). The QCRT identifies Criterion 4 Success Criteria, as a necessary PA component. According to the QCRT, PAs should include a rubric or other scoring tool such as a checklist or analytic rubric that are tightly aligned to the performance expectations of the intended learning outcomes. Success criteria are important because one intent of a tool such as this is to provide feedback to students about their work and how it can be improved. Kan and Bulut (2014) examined the effects of teacher experience and rubric use in PAs. Findings revealed that teachers who used rubrics to score PAs showed consistency among the scores given. In contrast, inconsistent scores resulted among teachers when a rubric was not used. Based upon my experiences in this study, I would like to add to the findings of Kan and Bulut to suggest that even when a rubric accompanies a PA, scoring outcomes may still be inconsistent and lack realization of what students know and are able to do if the rubric does not contain quality features, as recommended by the QCRT.

Of the 12 schools with performance assessments reviewed for quality in this study, eight were noted on Table 16 as follows: one school had no rubric included at all; six schools revealed less than limited evidence (≤.8); and two schools revealed limited evidence (1). One school showed barely above limited evidence (1.2) while another school revealed partial evidence (2), and one school began to approach full evidence (2.2). In stark contrast to the quantitative data, qualitative data gathered during focus groups revealed that 70% of participants specified the use of rubrics or checklists to ensure students understood the success criteria of a PA. In other words, while participants stated that success criteria were provided to help students understand the expectations, the data suggested that success criteria were weak in substance and did not clearly account for the conditions of the assessments.

Next, as two of the lowest rated criteria on the QCRT, feasibility and accessibility are examined as complementary.

Criterion 6: Accessibility; and Criterion 7: Feasibility

To address accessibility and feasibility, I begin with a discussion regarding the feasibility of a PA because a lack of this quality in a performance assessment may weaken performance assessment accessibility for students. Feasibility is addressed in the QCRT in several ways. A quality performance assessment should include student-facing prompts, directions, resources/materials and scoring tools. As well, resources and materials that students need to complete the PA should be realistic and easily accessible. The duration of time it takes to complete a PA should be clearly indicated and aligned to the complexity of the task and, if the PA is implemented over multiple lessons, there is a schedule. As well, connections to students' prior learning should fit within the scope of the learning sequence of the PA. As is stated in the data of this study, the lack of feasibility, as seen in Criterion 7, was near the lowest rating (.9) on

the QCRT rubric. A minimal amount of PAs provided student directions, prompts, connections to prior learning, or scoring tools. Quality data from eight schools reported below limited evidence (≤.7) with one school at limited evidence (1). Two schools approached partial evidence (≤1.7), and one school approached full evidence (2.5). Evidence from focus group question three, Sub-question C indicated that each school provided ample amounts of clearly communicated procedures to students regarding the implementation of PAs.

Providing examples was noted by 60% of schools while 70% of schools indicated they furnished a rubric or checklist to communicate procedures. Written directions and repeating step-by-step directions were mentioned by 50% of schools and finally, teacher feedback was mentioned by 40% of schools as a way to communicate expected procedures. As seen above, there is clear inconsistency between what the quantitative data revealed regarding weak quality ratings for feasibility and what teachers stated they actually did to provide student directions, prompts, connections to prior learning and scoring tools. PAs that lack feasibility can cause frustration for students because they may have to spend unnecessary time and energy trying to figure out exactly what is expected in order to be successful. As well, teachers may lose instructional time when they must explain and re-explain the expectations. Addressing feasibility concisely before implementing a PA can set students up for success and save valuable time.

When a PA is not feasible it cannot accommodate the participation of all students, which speaks to Criterion 6, accessibility. In this study, accessibility received an overall quality rating of 1, revealing limited evidence of quality and appearing as next to the lowest overall rating on the QCRT. Every PA should have clear directions for teachers identifying appropriate supports or alternatives to facilitate accessibility while maintaining validity (does it measure what it is supposed to measure) and reliability (does it *consistently* measure what it is supposed to

measure). Also, a PA should allow for differentiating the ways that students demonstrate their knowledge.

A similar disconnect between quantitative and qualitative data appeared with criterion 6 as seen with criterion 4 and criterion 7. Shown in Table 19**,** 100% of schools mentioned that the resources/materials needed for students to be able to implement a PA were provided by the teacher in order to make the PA accessible for all students. However, Table 20 indicates that only 70% of participants actually did provide rubrics/checklists to communicate procedures of a PA with students. Providing examples was only mentioned by 60% of schools while written directions and repeating step-by-step directions were cited by 50% of schools. And finally, only 40% of schools stated that teacher feedback was provided as a way to communicate procedures. While teachers claimed they made PAs accessible for all students, the actual quantitative data reflected otherwise. Authenticity is addressed in the next section and is the highest rated criteria in the QCRT.

Criterion 2: Authenticity

With the highest overall rated criteria (2.3) in the QCRT data, authenticity was seen as a relative strength among all criteria. Oberg (2010) and Koh et al., (2011) identified authentic or real-world contexts as one of the most significant characteristics of quality PAs. Based on the outcomes seen in the qualitative data of this study, 60% of schools participating in focus groups referenced their experiences in division-wide PBL training as contributing to their preparation to design or select and implement PAs. Since the main tenet of PBL is for students to learn by actively engaging in real-world or personally meaningful projects, strong alignment of PAs to authenticity was not surprising ("What is PBL?", n.d.). Allowing students the autonomy to choose the topics that interest and challenge them was noted by school 2 focus group as bringing

about the most original and genuine assessment results. As well, teachers in the School 4 focus group mentioned that students naturally gravitate toward topics they feel strongly about when given voice and choice. Authenticity also connects to the work of a discipline, such as history (Curry & Smith, 2017). For example, teaching students to research and write about the past is a real-world task of historians.

Document-based Questions (DBQs) are PAs that connect to the real-world tasks of historians and were mentioned by 60% of schools stating that participation in DBQ professional learning workshops enhanced their preparation to implement quality PAs. Teachers in School 7 focus group referred to DBQs as a way for students to take historical documents and flesh them out in order to gain a clearer understanding of history. While DBQs may not offer as much voice and choice as PBL, they do ask students to provide evidence to back up a claim, which is a skill that can be connected to real-world, workplace experiences. In support of this finding, Gareis (2018) pointed out that as a means of gathering evidence of student learning, PAs require students to think at higher cognitive levels, undertake a skill-based process and/or product, and are authentic to the discipline or real world. Similarly, Curry and Smith (2017) acknowledged that while PAs can work to solve real-world problems, within the context of social studies, PAs can also be directly meaningful to students as future citizens.

While findings indicated that some schools implemented DBQs with students, DBQs were not submitted for a quality review in this study. DBQs are generally more prescriptive in nature and offer minimal student voice and choice. However, tasks similar to that of a DBQ appeared in some of the PAs submitted for review. For example, while most PAs submitted for review included student choice of topic and product, they also contained experiences similar to DBQ such as collaboration opportunities for students and instances of making evidence-based

claims. This may suggest that schools who had knowledge of DBQ instructional practices were able to plan their own PAs with DBQ design strategies in mind. In general, however, skills helpful for students in real-world jobs, like creating media-based presentations, were more often seen as products created by students rather than evidence-based essays, typically thought of as the real-work work of historians. Worthy of note, schools with higher authenticity ratings were mostly those with PBL and DBQ experience. Of the schools with the lowest ratings for authenticity, DBQs were not mentioned and two out of three did not participate in PBL training. Although authenticity received the overall highest rating (2.3) on the QCRT, it is important to note that a 2.3 is only slightly above partial evidence and thus, leaving room to improve. The next highest rated quality criteria referenced in the data was language use for expressing reasoning.

Criterion 3: Language Use for Expressing Reasoning

Language use for expressing reasoning was slightly above partial evidence (2.2) quality alignment and may be considered a relative strength as the second highest QCRT rating. When students express reasoning they communicate feelings, thoughts, ideas, and information. According to the QCRT, PAs should support students' language use by providing multiple means of accessing and using developmentally appropriate, academic, and disciplinary language, e.g., text, video, audio, and oral. Curry and Smith (2017) supported three important aspects of PAs that depend upon language use such as “demonstration of knowledge, ability to reason, and ability to communicate conclusions” (p. 169). The National Council for the Social Studies (2013) provided an example of an inquiry to engage a variety of language types in which students work together to create a civilization from the ground up, using technology for research, collaboration and dissemination. In this instance, students can practice the authentic use of language during

conversations with peers while utilizing functional, academic and disciplinary language. Language use for expressing reasoning is a skill often practiced during the PBL and DBQ implementations and must be recognized and included in professional development for teachers in this study to enhance or improve its practice. The two schools that showed full-evidence (3) of quality for criterion 3, along with the five schools that approached full-evidence (≤2.8), offered assorted ways for students to engage with academic and disciplinary language. There were opportunities to collaborate with peers through group work or during exhibit presentations. Schools with quality ratings below partial evidence (≤1.8) offered fewer choices and opportunities in the ways students communicated their reasoning and minimal chances to collaborate with classmates.

Next, Criterion 1, Standards/Intended Learning Outcomes approached a quality rating of partial evidence and will be discussed.

Criterion 1: Standards/Intended Learning Outcomes

Clear alignment to the Standards of Learning is a key quality feature of performance assessments. According to the QCRT, PAs should be developmentally appropriate for target students and aligned to the grade-level scope and sequence or grade-level curriculum. The overall quality rating for this criteria was fairly weak with a score that approached partial evidence (1.7). When teachers were asked to describe their typical practice of aligning a PA to the content that is taught, school 10 mentioned making sure the PA was clearly aligned to the SOL and to make sure students know this too. School 12 asked the question, "What is my end result?" as their go-to question to accomplish a goal. General feedback from schools concluded that typical practices of aligning PAs to content is to start with the end in mind; align to SOL; begin with the skills and standards; and backward design. While 80% of schools mentioned

alignment to SOL and 60% of schools referred to backward design, there was very little evidence of this alignment clearly listed in a task template, as described in criterion 1A. Due to my own background knowledge of the SOL, I was able to generally discern that most PAs did align to the SOL. The low ratings in criterion 1A came from the fact that most PAs did not appear in a structured template. Hence, the quality of criterion 1A would have been greatly improved if a structured template had been available to explicitly show alignment to grade level scope and sequence, resources/materials, and student products. The utility of a template is immeasurable because it sets up an equitable process for all teachers to deliberately think about and plan for the said alignment that all students need to be successful.

Additional topics included in Criterion 1 are whether PAs go beyond simple recall to elicit evidence of complex student thinking and opportunities for demonstrating deeper learning competencies. The overall quality rating for criterion 1 approached partial evidence (1.7). An approach to partial evidence reveals weaknesses in the ability to recognize and design PAs that support opportunities for students to engage with deeper learning competencies. Important to note is that during focus group discussions, 100% of schools mentioned the PLC as a source that contributed to the preparation to design (or select) and implement PAs. Thus, it may be suggested that the efficacy of PCPS PLCs to routinely develop and demonstrate PAs that elicit evidence of complex student thinking and deeper learning competencies is weak.

Next, Criterion 5 Student Resources, approached a quality rating of partial evidence (1.8) and will be discussed.

Criterion 5: Student Directions, Prompts, and Resources/Materials

When trying to determine the quality of feasibility and accessibility of a PA, Criterion 5 provides some of the basis for this determination. There was 100% consensus among schools in

focus groups with respect to providing the essential resources for all students when implementing performance assessments. However, quantitative data showed that the quality of student directions, prompts, and resources/materials was weak with only an approach to partial evidence (1.8). During focus groups, teachers mainly spoke about making sure that resources/materials were available for students in order to provide for equitable opportunities for success. As well, schools mainly mentioned verbal directions as one of the common ways to help students understand what they needed to do for a PA, as opposed to clearly written directions. Students are less likely to be successful when they do not have a clear PA prompt or directions to understand the purpose of a task and the intended learning outcomes being assessed. The next section will provide suggested implications of this study for professional practice and introduce the *History and Social Science Standards of Learning Skills Progression Chart* (VDOE, 2015) and the *History and Social Science State Developed Common Rubrics* (VDOE, 2020) to support said practice.

Implications for Practice

Professional development for middle school USI and USII social studies teachers in PCPS aimed at recognizing, designing, and implementing quality performance assessments would help to improve the quality of these assessments. As the curriculum specialist in PCPS, I am responsible for leading the professional development for teachers and using the results of the data from this study to guide my planning. Regarding professional development, Guskey and Yoon (2009) stated that many education leaders criticize ineffective workshops as a waste of time and money, especially the one-shot variety that offers no genuine follow-up or sustained support. As well, Koh et al., (2011) revealed findings in a 2-year study that ongoing and sustained professional development in designing and implementing authentic assessments and

rubrics was more effective than ad-hoc, 1– or 2–day workshops to build teachers' capacity in improving professional practice. The professional development I design will be ongoing through the school year and grounded in the findings of this study. My desire is to build capacity so that teachers are now and in the future able to recognize and implement quality performance assessments.

Professional development will begin by meeting individually with department chairpersons at each middle school to model the exact workshop that I will deliver to their teachers. My intent is to prepare department leaders ahead of time so that they can support their teachers after my workshop. The workshop will include slides with notes and links to resources that I will share with teachers to keep for future reference. Teachers will be expected to come to the workshop with a sample performance assessment. Once the initial workshop content slides are presented, teachers will review their own PAs for quality to decide what revisions need to be made based upon what they learned. I will then schedule another school visit to meet with teachers to review their revised PAs and provide support as needed. The next few paragraphs described additional details about the workshops. The *History and Social Science Standards of Learning Skills Progression Chart*, launched in 2015, and *VDOE Middle School Common Rubric* in 2020, will be introduced as important tools to support quality performance assessments.

In 2015, the VDOE identified the grade/course at which specific skills were formally introduced in the History and Social Science SOL and made available in the *History and Social Science Standards of Learning Skills Progression Chart*. On the chart, the darkest shade of pink denotes skills that students should have already mastered. The lighter pinks denote skills that are introduced or being scaffolded in the course. During the workshop, teachers will use the VDOE Skills Progression Chart by finding their course listed at the top of the chart. Teachers will make

note of the lighter shades of pink for their course as the skills that should be introduced or scaffolded for their course. Figure 29 (see Appendix P) shows the 10 skills on the progression chart by grade or course.

After teachers identify the light pink skills on the progression chart they will compare them to the skills shown on the VDOE common rubric for middle school social studies. Noted at the top of the middle school rubric, as seen in Appendix Q, are the core expectations for each performance assessment. The VDOE expectation for social studies performance assessments is that they consistently assess .1a and .1c skills using this rubric in an effort to show growth over time. These skills encompass the following: accuracy of content, synthesizing information sources, and explaining evidence and align to the skills progression chart. Teachers should then choose between one and three additional skills to be included in the performance assessment from the task specific concepts and skills section of the rubric that also align to the skills progression chart, as seen in Appendix R.

Teachers will also be introduced to a PA template that is closely aligned to the QCRT and a sample PA developed using the template as a model to guide teachers with their own PA. Finally, teachers will consider what they have learned during the workshop to apply to their own PAs by making updates and aligning to the template. Within 3-4 weeks I will meet with teachers individually to review their updated PAs for alignment to the QCRT. Once PAs have been updated to align to the QCRT, they will be added to an electronic portal for all social studies teachers to access as a bank of resources.

I will continue to follow-up with department chairs and teachers throughout the school year, and into the next, to review the quality of teacher-selected performance assessments. Providing coaching support in future years will help maintain the development of quality

teacher-selected performance assessments and build the capacity of current teachers to train new teachers as they are introduced to the idea of maintaining quality PAs. The review of performance assessments for quality should be an ongoing practice in order to best serve students now and in the future.

The relationship between professional development and improvements in student learning can be complex. Workshops of the one-time variety with no follow-up do not sustain best instructional practices. However, well-planned workshops that include research-based practices and active learning for participants with opportunities to adapt the practices to the unique classrooms of teachers may indeed support student learning (Guskey & Yoon, 2009). The workshops I implement will attempt to follow best practice based upon the research practices mentioned by Guskey and Yoon in their 2009 research synthesis, *What Works in Professional Development?* Guskey and Yoon (2009) also pointed out that there are few studies conducted on the impact of professional development on specific student achievement gains.

Recommendations for Future Research

School districts in Virginia continue to transition to a more balanced approach to instruction and assessment. Research can be conducted to support this transition and can be made as a result of this study and in more than one subject area. A methodological choice of a randomized experimental design is suggested in order to establish a cause-and-effect relationship between quality performance assessments and improved pedagogical practices (Mertens & Wilson, 2012). Another cause-and-effect relationship to consider as a research topic is the relationship between quality performance assessments and student learning. In this study, the independent variable would be the performance assessment reviewed for quality features and revised based upon the QCRT and the original performance assessment. The group of students

receiving the quality performance assessment would be the experimental group; the group that receives the original PA would be the control group. The expected dependent variables would be improvements in the way students think critically, mastering rigorous academic content, making authentic or real-world connections, and generally changing the way students engage in deeper learning.

The QCRT purports to measure the application of content knowledge and skills and to support comparability in rigor and quality across the state. Therefore, a research study designed to test that claim may begin with the question: To what degree does the QCRT support comparability in rigor and quality across the state? The evaluation could use a theory-based approach where the outcome would be ratings that show the relationship between the variables in the QCRT and comparable outcomes across the state. Data collection would consist of a review of performance assessments using the QCRT by each school division. A comparison of the data among school divisions would show to what degree the scores align. Qualitative data could be collected and compared from the written interpretations of rater rationales used to qualify criteria as well as focus groups among school divisions.

Conclusion

Quality performance assessments have the potential to promote deeper learning opportunities for students and were valued by PCPS middle school social studies teachers. A significant number of teachers communicated great support for their use because PAs provide more opportunities for purposeful differentiation among learners. Some teachers, however, indicated frustration that PAs take more class time but also recognized that they are a more equitable way to assess because students are often able to choose how they will demonstrate what they know. Teachers also mentioned that they feel empowered and valued as professionals

because they have the autonomy and freedom to design and implement their own PAs based upon the strengths and weaknesses of their students. They also appreciated the opportunity to collaborate with their colleagues when choosing PAs and the chance to recognize that student weaknesses soon become strengths with teacher feedback.

While there is positivity surrounding the use of PAs by teachers, the quantitative data analysis findings regarding the degree of quality seen in teacher-selected PAs was weak (limited to partial evidence). I found that teachers were very familiar with how to develop projects and based on this study, some students are getting the opportunity to participate in work authentic to the discipline or in developing skills authentic to the real world of future careers. However, what was not evident was the extent to which students were actually engaging in deeper learning projects accompanied by clear directions and rubrics that outlined what all students needed to do to be successful. To improve the quality of PAs in PCPS, the design of PAs need to accommodate the participation of all students to include clear student-facing prompts, directions, and scoring tools that align to deeper learning experiences.

References

Adeyemi, B. (2015). The efficacy of authentic assessment and portfolio assessment in the learning of social studies in junior secondary schools in Osun State, Nigeria. *Ife PsychologyIA, 23*(2), 125-132.

Archbald, D. & Newman, F. (1988). *Beyond standardized testing: assessing authentic Academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals Assessment. https://eric.ed.gov/?id=ED301587

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267. https://eric.ed.gov/?id=EJ782544

Barber, W., King, S. & Buchanan, S. (2015). Problem based learning and authentic assessment in digital pedagogy: Embracing the role of collaborative communities. *The Electronic Journal of e-Learning, 13*(2), 59-67. http://www.ejel.org

Bennet, L. L. & Berson, M. J. (Eds.). (2007). Digital age: Technology-based k-12 lesson plans for social studies. *NCSS Bulletin,* 105.

Bland, L. & Gareis, C. (2018). Performance assessments: A review of definitions, quality, characteristics, and outcomes associated with their use in K-12 schools. *The Teacher Educators' Journal, 11*, 52-69. http://files.umwblogs.org/sites/2385/2009/05/18204847/The-Teacher-Educators-Journal-spring-2018.pdf

Brookhart, S. (2013). The public understanding of assessment in educational reform in the United States. *Oxford Review of Education, 39,* 52-71. https://eric.ed.gov/?id=EJ1012332

Buck Institute for Education. (n.d.). *What is PBL?* https://www.pblworks.org/what-is-pbl

Capistrano Unified School District: 2017-18 DBQ Initiative, Grades 6-12.

Chang, C., Tseng, K., & Liang, C. (2014). Is reflection performance correlated to the learning effect in a web-based portfolio assessment environment for middle school students? *Asia-Pacific Education Researcher 23*(1), 73-82. DOI 10.1007/s40299-013-0087-8

Comer, M. (2011). Young adult literature and alternative assessment measures. *Theory Into Practice 50,* 239-246. https://eric.ed.gov/?id=EJ932961

Corcoran, C., Dershimer, E., & Tichenor, M. (2004). A teacher's guide to alternative assessment: Taking the first steps. *The Clearing House 77*(5), 213-216. http://www.jstor.org/stable/30189900

Curry, K. & Smith, D. (2017). Assessment practices in social studies classrooms: Results from a longitudinal survey. *Social Studies Research and Practice 12*(2), 168-181. http://www.emeraldinsight.com/1933-5415.htm

Darling-Hammond, L. & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning.* Stanford University, Stanford Center for Opportunity Policy in Education. https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-executive-summary_0.pdf

Duffy, M., Jones, J., & Thomas, S. (1999). Using portfolios to foster independent thinking. *Intervention in School and Clinic 35*(1), 34-37. https://doi.org/cpx2fw

Dunbar, S., Koretz, D., & Hoover, H. (1991). Quality control in the development and use of performance assessments. *Applied Measurement In Education 4*(4), 289-303. https://www.tandfonline.com/doi/abs/10.1207/s15324818ame0404_3

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

Frey, B. & Schmitt, V. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal, 5*(3), 107-117. https://eric.ed.gov/?id=EJ901788

Frey, B., Schmitt, V. & Allen, J. (2012). Defining authentic classroom assessment. *Practical Assessment Research & Evaluation, 17*(2), 1-18. https://scholarworks.umass.edu/pare/vol17/iss1/2/

Gallavan, N., & Kottler, E. (2009). Constructing rubrics and assessing progress collaboratively with social studies students. *The Social Studies, 100*(4), 154-158. https://eric.ed.gov/?id=EJ846815

Gareis, C., (2017, March), *PBA Development* ["PowerPoint slides"]. School of Education, College of William and Mary.

Guskey, T., & Yoon, S. (2009, March). What works in professional development. *Phi Delta Kappan Magazine,* 495- 502.

Hallam, R., Grisham-Brown, J., Gao, X., & Brookshire, R. R. (2007). The effects of outcomes-driven authentic assessment on classroom quality. *Early Childhood Research & Practice, 9*(2). http://go.galegroup.com/ps/i.do?id=GALE%7CA1787 16829&v=2.1&u=PCPShigh&it=r&p=GPS&sw=w

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Johnson, K. H. (2016). *A program evaluation of the document-based question project*. https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?referer= https://www.google.com/&httpsredir=1&article=3710&context=dissertations

Johnson, K. H. (2016). A program evaluation of the document-based question project. *Walden Dissertations and Doctoral Studies. 2607*. https://scholoarworks.waldenu.edu/dissertations/2607

Kan, A., & Bulut, O. (2014). Crossed random-effect modeling: Examining the effects of teacher experience and rubric use in performance assessments. *Eurasian Journal of Educational Research, 57*, 1-28. https://files.eric.ed.gov/fulltext/EJ1056261.pdf

Klein, A. (2016, September 27). ESSA clears out underbrush on school improvement path. *Education Week*. http://www.edweek.org /ew/articles/2016/09/28/ essa-clears-out-underbrush-on-school-improvement.html

Kloser, M., Borko, H., Martinex, J., Stecher, B., & Luskin, R. (2017). Evidence of middle school science assessment practice from classroom-based portfolios. *Science Education, 101*(2), 209-231.

Koh, K., Tan, C., & Ng, P. (2011). Creating thinking schools through authentic assessment: The case in Singapore. *Educational Assessment Evaluation and Accountability 24*, 135-149, doi: 10.1007/s11092-011-9138-y

Lai, E., (2011, May). Performance-based assessment: Some new thoughts on an old idea: 20th Issue. Pearson Education, Inc. http://www.pearsonassessments.com

Marzano, R., (2003). *What works in schools translating research into action*. Association for Supervision and Curriculum Development. https://eric.ed.gov/?id=ED474629

Mertens, D. & Wilson, A. (2012). *Program evaluation theory and practice a comprehensive guide*. The Guilford Press.

Moeller, A., Theiler, J. & Chaorong, W., (2012). Goal setting and student achievement: A longitudinal study. *The Modern Language Journal 96*(2), 153-169. https://www.jstor.org/stable/41684067

Monte-Sano, C., (2008). Qualities of historical writing instruction: A comparative case study of two teachers' practices. *American Educational Research Journal, 45*(4), 1045-1079. https://doi.org/10.3102/0002831208319733

Moss, C. & Brookhart, S., (2008, November). *Formative assessment that empowers.* ASCD. https://www.ascd.org/el/articles/formative-assessment-that-empowers

Mueller, J., (2018). *Authentic assessment toolbox*. North Central College. http://www/jfmueller.faculty.noctrl.edu/toolbox/index.htm

National Council for the Social Studies (NCSS) (2013). *The college, career, and civic life (C3) framework for social studies state standards: Guidance for enhancing the rigor of K-12 civics, economics, geography, and history*. NCSS.

Oberg, C. (2010). Guiding classroom instruction through performance assessment. *Journal of Case Studies in Accreditation and Assessment, 1*, 1-11. https://files.eric.ed.gov/fulltext/EJ1055507.pdf

Pang, V.O., et.al., (Eds.). (2010). The human impact of natural disasters; issues for the inquiry-based classroom. *NCSS Bulletin* 110, 113 pp., 2010.

Project Appleseed. (2014). *Testing, testing* [Brochure]. http://www.projectappleseed.org/assessment

Stanford Center for Assessment, Learning and Equity. (n.d.). *About SCALE.* https://scale.stanford.edu/about

Stake, R. E. (2004). *Standard-based and responsive evaluation*. Sage.

Staples, S. (2014a). *Clarifications regarding 2014 legislation impacting the standards of learning testing program.* Virginia Department of Education. http://www.doe.virginia.gov/ administrators/superintendents _memos/2014/216-14.shtml

Staples, S. (2014b). *Local alternative assessment guidelines approved by the Board of Education*. Virginia Department of Education. http://www.doe.virginia/gov/administrations/superintendents_memos/2014/278-14.shtml

Staples, S. (2016a, November). *Local alternative assessment guidelines for 2016-2017 through 2018-2019* (Attachment A to Supt's Memo No. 284-16). Virginia Department of Education.

Staples, S. (2016b). *Virginia's public comments on U.S. department of education proposed regulations for accountability and state plan regulations*. Virginia Department of Education. http://www.doe.virginia.gov/ federal_programs /esea/essa/ communications /index.shtml

Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability.* Stanford Center for Opportunity Policy in Education.

Stufflebeam, D. & Shinkfield, A. (2007). *Evaluation, theory, models, and applications.* Jossey-Bass.

VanTassel-Baska, J. (2013). Performance-based assessment the road to authentic learning for the gifted. *Gifted Child Today*, *37*(1), 41-47. https://doi.org/10.1177/1076217513509618

Virginia Department of Education. (2014a). *Assessment literacy glossary.* http://www.doe.virginia.gov/testing/local_assessments/glossary.pdf

Virginia Department of Education. (2014b). *Framework for local alternative assessment guidelines.* http://www.doe.virginia.gov/testing/local_ assessments/framework-for-laa-implementation.pdf

Virginia Department of Education. (2015). *History and social science standards of learning skills progression chart.* Richmond, VA.

Virginia Department of Education. (2016a). *Profile of a Virginia graduate.* Richmond, VA.

Virginia Department of Education. (2016b, November). *Local alternative assessment guidelines for 2016-2017 through 2018-2019* (Attachment A to Supt's Memo No. 284-16). Steven Staples.

Virginia Department of Education. (2018). *Virginia quality criterion tool for performance-based assessments*.

Virginia Department of Education. (2019a, January). *Virginia quality criterion tool for performance-based assessments*. Richmond, VA.

Virginia Department of Education. (2019b, January). *Guidelines for Local Alternative Assessments for 2018-2019 through 2019-2020* (Attachment A to Supt's Memo No. 025-19). Richmond, VA: James Lane.

Virginia Department of Education. (2020). *History and social science state developed rubrics.*

Virginia Department of Education. (2021). *Guidelines for local alternative assessments for 2021-2022 and Beyond* (Attachment A to Supt's Memo No. #214-21).

APPENDIX A

Portfolios: Types, Purposes, and Samples of Work

Growth Portfolio	
Purpose	**Sample Work**
a. to show growth or change over time	● Early and later pieces of work ● Early and later tests/scores ● Rough drafts and final drafts ● Reflections on growth ● Goal-setting sheets ● Reflections on progress toward goal(s)
b. to help develop process skills	● Samples which reflect growth of process ● Self-reflection sheets accompanying samples of work ● Reflection sheets from teacher or peer ● Identification of strengths/weaknesses ● Goal-setting sheets ● Reflections on progress towards goal(s)
c. to identify strengths and weaknesses	● Samples of work reflecting specifically identified strengths and weaknesses ● Reflections on strengths and weaknesses of samples ● Goal-setting sheets ● Reflection on progress towards goal(s)
d. to track development of one or more products or performances	● Obviously, drafts of the specific product or performance to be tracked ● Self-reflection on drafts ● Reflection sheets from teacher or peer
Showcase Portfolio	
Purpose	**Sample Work**
a. to showcase end-of-year/semester accomplishments	● Samples of best work ● Sample of earlier and later work to document progress ● Final tests or scores ● Discussion of growth over semester/year ● Awards or other recognition ● Teacher or peer comments
b. to prepare a sample of best work for employment or college admission	● Cover letter ● Sample of work ● Reflection on process of creating sample of work ● Reflection on growth ● Teacher or peer comments ● Description of knowledge/skills work indicates
c. to showcase student perceptions of favorite, best or most important	● Samples of student's favorite, best or most important work ● Drafts of that work to illustrate path taken to its final form ● Commentary on strengths/weaknesses of work ● Reflection on why it is favorite, best or most important ● Reflection on what has been learned from work ● Teacher or peer comments
d. to communicate a student's current aptitude	● Representative sample of current work ● Match of work with standards accomplished ● Self-reflection on current aptitudes ● Teacher reflection on student's aptitudes

	● Identification of future goals
Evaluation Portfolio	
Purpose	**Sample Work**
a. to document achievement for grading	● Samples of representative work in each subject/unit/topic to be graded ● Samples of work documenting level of achievement on course/grade-level goals/standards/objectives ● Tests/scores ● Rubrics/criteria used for evaluation of work (when applied) ● Self-reflection on how well samples indicate attainment of course/grade-level goal/standards/objectives ● Teacher reflection on attainment of goals/standards ● Identification of strengths/weaknesses
b. to document progress towards standards	● List of applicable goals and standards ● Representative samples of work aligned with respective goals/standards ● Rubrics/criteria used for evaluation of work ● Self-reflection on how well samples indicate attainment of course/grade-level goals/standards/objectives ● Teacher reflection of attainment of goals/standards ● Analysis or evidence of progress made toward standards over course of semester/year
c. to place students appropriately	● Representative samples of current work ● Representative samples of earlier work to indicate rate of progress ● Classroom tests/scores ● External tests/evaluation ● Match of work with standards accomplished ● Self-reflection on current aptitudes ● Teacher reflection on student's aptitudes ● Parent reflection on student's aptitudes ● Other professionals' reflection on student's aptitudes

Note. From *Authentic assessment toolbox*, J. Mueller, 2018, North Central College. http://www/jfmueller.faculty.noctrl.edu/toolbox/index.htm

APPENDIX B

GUIDELINES FOR LOCAL ALTERNATIVE ASSESSMENTS

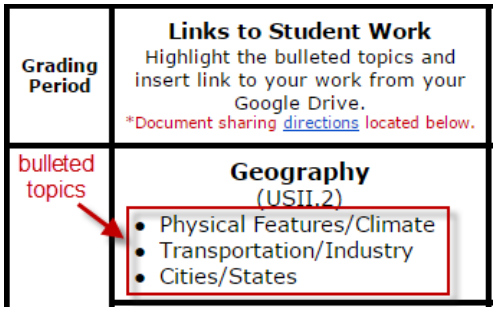
SOCIAL STUDIES PORTFOLIO ASSESSMENTS 2017-18 SCHOOL YEAR

Purpose: The Virginia Department of Education has required alternative assessments be provided for courses with eliminated SOL tests to verify the content is being taught. To comply with this state guideline, PCPS will require a grade 3, USI, and USII ***snapshot*** social studies portfolio for each student representing the school year 2017-18 in each of the four strands: History, Geography, Economics, and Civics.

Social Studies Objectives:

Students will:

1. save the course appropriate portfolio file in their Google Drive
2. link no less than two or no more than four assignments that represent their ***best*** work based on the bulleted topics under each category (only one assignment can be linked to each bulleted topic)



3. self-assess using the 4-3-2-1 rating scale and peer review throughout the school year

Information:

- Teachers have flexibility in deciding how portfolios can be set up as they work within the requirements listed in objectives 1-3.
- Students should have opportunities to practice essential skills located in columns of charts throughout the curriculum framework. Evidence of skills-based instruction should be evident in portfolio assignments.
- Teachers are required to make sure students keep up with adding materials to the portfolio during the school year.
- Student artifact examples may include, but are not limited to:
    - Interactive Notebook
    - Reflective Journal
    - Content Writing Prompt
    - Common Assessment
    - Performance Based Assignment
    - PBL Project
- Self-assessment and peer review should occur one time each semester.
    - Students will use a 4-3-2-1 rating scale for self-assessments that will become a part of the portfolio.
- Teachers should monitor student portfolios throughout the year and provide documentation of this monitoring in their lesson plans.

APPENDIX C

VIRGINIA QUALITY CRITERIA REVIEW TOOL FOR PERFORMANCE ASSESSMENTS

Virginia Quality Criteria Review Tool for Performance Assessments

Revised: June 2019

This document details a set of criteria for the development of performance assessments that measure the application of content knowledge and skills. The criteria are designed to support comparability in rigor and quality across the state.

Briefly, the seven criteria include:

1: Standards/Intended Learning Outcomes
2: Authenticity
3: Language Use for Expressing Reasoning
4: Success Criteria for Students
5: Student Directions, Prompt, and Resources/Materials
6: Accessibility
7: Feasibility

It is not expected that every quality performance assessment will have full evidence that corresponds to each description within the Virginia Quality Criteria Review Tool. For example, it is not expected that a task include each of the deeper learning competencies listed in criterion 1C (i.e., mastering rigorous academic content; learning how to think critically and solve problems; working collaboratively; communicating effectively; directing one's own learning; and developing an academic mindset), but a quality task should provide the opportunity for students to demonstrate one or more of these competencies. Additionally, the criteria may be considered in any order that suits the division's needs and purpose. School divisions are encouraged to use this tool not only as they develop new performance assessments, but also as they evaluate the quality of existing tasks. If a task does not fully meet the expectations detailed for that criterion, the school division should use the evidence or rationale to modify and improve the task before using that task with students.

Virginia Quality Criteria Review Tool for Performance Assessments

Revised: June 2019

Criterion 1: Standards/Intended Learning Outcomes

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
1A	Virginia Standards of Learning selected for the performance assessment are clearly listed in a task template, developmentally appropriate for target students, and aligned to the grade-level scope and sequence or grade-level curriculum. Performance assessment components, resources/materials, and student products are aligned to the listed SOLs.		
1B	The performance assessment goes beyond simple recall, elicits evidence of complex student thinking, and requires application of disciplinary or cross-disciplinary concepts, practices, and/or transferable skills, such as application, analysis, evaluation, synthesis, or original creation.		
1C	The performance assessment provides an opportunity for students to develop and demonstrate (even if not explicitly assessed): • Deeper learning competencies, defined as mastering rigorous academic content; learning how to think critically and solve problems; working collaboratively; communicating effectively; directing one's own learning; and developing an academic mindset. The performance assessment may also provide opportunities for students to develop and demonstrate: • Life-Ready competencies defined by the Profile of a Virginia Graduate as content knowledge, career planning, workplace skills, and community and		

#	Description	Quality Rating	Evidence or Rationale
	civic responsibility; • Technology-related competencies; • Integration of intended learning outcomes from two or more subjects.		

Criterion 2: Authenticity

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
2	The performance assessment is authentic along the dimensions: • The performance assessment's topic, context (scenario), materials/resources, products, and purpose/audience (i.e., what students are asked to do and for whom) are relevant to the real-world, students' community, students' interests, future careers, or other meaningful context. • The performance assessment asks students to do work authentic to the discipline (i.e., what adult practitioners of the discipline do), such as science inquiry; math problem-solving; analyzing and critiquing a text; analyzing and evaluating historical sources.		

Criterion 3: Language Use for Expressing Reasoning

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
3A	The performance assessment supports language use and development by providing multiple means of accessing and using developmentally appropriate academic and disciplinary language for the students to express their		

#	Description	Quality Rating	Evidence or Rationale
	reasoning.		
3B	The performance assessment should require students to use one or more forms of language to communicate their reasoning. The performance assessment may provide access to functional, academic, and disciplinary language in various forms of language media (text, video, audio, oral) OR provide opportunity to practice the use of language through multiple means of expression and language production (text, language media production, oral language, or conversation with peers).		

Criterion 4: Success Criteria for Students

The Virginia Department of Education's Common Rubrics, when available, should be used to evaluate and score student work. The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
4A	The performance assessment includes a rubric or other appropriate scoring tools (e.g., checklist, analytic rubric) with scoring dimensions that are tightly aligned to performance expectations of the intended learning outcomes targeted within the performance assessment. Criteria should include language objectives, if applicable.		
4B	The scoring tool is written clearly and concisely, with audience-friendly language, as appropriate. Language of the scoring tool should describe how a response demonstrates performance expectations so that the tool may be used to provide feedback to students about their work and how it can be improved.		
4C	The scoring tool or feedback methodology should be used across performance assessments within the course so that results on the performance assessment can be used to communicate a consistent set of expectations to students, monitor students' academic growth over time, inform		

#	Description	Quality Rating	Evidence or Rationale
	instructional decisions, and communicate student proficiency to others (e.g., parents/guardians).		

Criterion 5: Student Directions, Prompt, and Resources/Materials

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
5A	The student-facing task prompt, directions, and resources/materials are aligned to the intended learning outcomes, task purpose, and the performance expectations being assessed (i.e., the student product will provide evidence of the performance expectations).		
5B	The student-facing task prompt, directions, and resources/materials are clear, complete, written in accessible language appropriate to the grade level, and organized for students in an accessible format.		
5C	The task prompt/directions, topic, context (scenario), and materials/resources are sensitive to the community and free of bias.		

Criterion 6: Accessibility

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
6A	The performance assessment is designed to accommodate the participation of all students. Directions for teachers for the performance assessment identify appropriate supports or alternatives to facilitate accessibility while maintaining the validity and reliability of the assessment.		
6B	The performance assessment is accessible and allows for differentiating the ways that students demonstrate their knowledge such as through the application of principles of		

#	Description	Quality Rating	Evidence or Rationale
	Universal Design for Learning (UDL). Refer to the National Center on UDL at the Center for Applied Special Technology (CAST).		

Criterion 7: Feasibility

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
7A	Student-facing prompts, directions, resources/materials, and scoring tools are included. Resources and materials required by the performance assessment are realistic and easily accessible to teachers.		
7B	Duration of implementation of the performance assessment is indicated and is realistic for the complexity of the assessment and the scope of performance expectations being assessed.		
7C	If the performance assessment is implemented over multiple lessons, a schedule indicating how the performance assessment is implemented across the lessons is included. Information about students' prior learning and how the performance assessment fits within a learning sequence is included.		

	required by the performance assessment are realistic and easily accessible to teachers.		
7B	Duration of implementation of the performance assessment is: • indicated • realistic for the complexity of the assessment and the scope of performance expectations being assessed.		
7C	If the performance assessment is implemented over multiple lessons: • a schedule indicating how the performance assessment is implemented across the lessons is included • information about students' prior learning and how the performance assessment fits within a learning sequence is included.		

APPENDIX D

STUDENT PORTFOLIO FOR UNITED STATES HISTORY to 1865

2017-2018 United States History to 1865 Student Portfolio

Student Name:

Teacher:

My social studies goal for the year is to learn about...

(type here)

Self-Rating Scale: **4**-I totally could teach it to a friend. **3**-I get it. **2**-I am starting to get it. **1**-I still need a little help.

Grading Period	Links to Student Work *Highlight the bulleted topics and insert link to your work from your Google Drive.* **Document sharing directions located below.*	Student Self-Reflection Self-reflection helps you make sense of what you have learned. (Now I know...)	Self-Rating (4-3-2-1)
1st	**Geography** (USI.2) • Continents and Oceans • US Geographic Regions • US Water Features • Map and Globe Features		
	Pre-Columbian Times (USI.3) • Archaeology • American Indian Cultures • American Indian Resources		
2nd	**European Exploration** (USI.4) • Motivations • Interactions • West African Societies		
	Colonization & Settlement (USI.5) • Religious & Economic Events • Environment & Economics • Different Perspectives • Political & Economic Relationships		
3rd	**Revolution** (USI.6) • Colonial Dissatisfaction • Political Ideas • People & Events • Defeat of Great Britain		
	New Nation (USI.7) • Articles of Confederation • US Constitution • Presidents		
4th	**Expansion** (USI.8) • Western Territories • Westward Movement		
	Civil War (USI.9) • Issues • States Rights • Secession • Leaders • Events • Perspectives		

This year in social studies I learned...

(type here)

APPENDIX E

STUDENT PORTFOLIO FOR UNITED STATES HISTORY

from 1865 – PRESENT

2017-2018 United States History from 1865 to Present Student Portfolio

Student Name:

Teacher

My social studies goal for the year is to learn about...

(type here)

Self-Rating Scale: **4**-I totally could teach it to a friend. **3**-I get it. **2**-I am starting to get it. **1**-I still need a little help.

Grading Period	Links to Student Work Highlight the bulleted topics and insert link to your work from your Google Drive. *Document sharing directions located below.	Student Self-Reflection Self-reflection helps you make sense of what you have learned. (Now I know...)	Self-Rating (4-3-2-1)
1st	**Geography** (USII.2) • Physical Features/Climate • Transportation/Industry • Cities/States	Now I know... Now I know...	
	Reform (USII.3) • 13th, 14th, & 15th Amendments • Reconstruction • Leaders	Now I know... Now I know...	
	Development of Industry (USII.4abc) • Westward Expansion • Immigration • Segregation	Now I know... Now I know...	
2nd	**Development of Industry** (USII.4de) • Inventions/Industry	Now I know... Now I know...	
	• Progressive Movement		
	Early 20th Century (USII.5) • Spanish American War • Foreign Policy • World War I	Now I know... Now I know...	
3rd	**1920s & 1930s** (USII.6) • Technological Changes • Social & Economical Changes • Art/Music/Literature • The Great Depression	Now I know... Now I know...	
	World War II (USII.7) • Causes • Europe & the Pacific • Home Front	Now I know... Now I know...	
	Postwar United States (USII.8ab) • Rebuilding • Economy	Now I know... Now I know...	
4th	**Postwar United States** (USII.8cde) • The Cold War • Changing Society • Globalization	Now I know... Now I know...	
	Contemporary United States (USII.9) • Civil Rights • Technologies • People • Foreign Policy	Now I know... Now I know...	

This year in social studies I learned...

(type here)

APPENDIX F

FOCUS GROUP PARTICIPATION EMAIL REQUEST

Dear USI and USII Teachers,

You are invited to participate in one focus group conducted by Lynne Bland as part of her doctoral research study about the quality of performance assessments. The focus group discussions will last approximately 75 minutes and will concentrate on understanding the quality of performance assessments implemented with students as one artifact in an alternative portfolio assessment. You will be asked questions about your views on the quality of the performance assessments you have implemented with students and your views regarding other issues related to the topic.

Please understand that participation in the focus group is completely voluntary; that you may choose not to answer certain questions, and that you may withdraw and discontinue participation at any time with no negative consequences.

Further understand that your confidentiality will be protected at all times and that a fictitious name will be assigned to you after the focus group is complete, and that any identifying information/characteristics about you or your colleagues will be deleted. The transcripts and the tapes will be assigned a numerical code and or a pseudonym and kept in a password protected computer file. The tapes and transcripts created to document my responses will be destroyed after being interpreted and synthesized in writing as the final step of this study.

Please click here to sign up on a Google form if you are interested. I will contact you in the near future to schedule dates and times to meet.

Thank you for your consideration,

Lynne Bland

APPENDIX G

FOCUS GROUP QUESTIONS

Introductory Statement: The PCPS Alternative Portfolio Assessment was implemented in 2014 as a replacement to the SOL test for accountability for the USI and USII courses. As a part of a balanced approach to assessment, performance-based assessments are included as one existing artifact in the PCPS student alternative portfolio assessment. Teachers have the autonomy to choose or create the performance-based assessment to implement with students. Based upon your experiences with the alternative assessment, please answer the following questions:

1. Do you support the move toward the use of performance assessments to measure student learning? Why do you feel this way?
2. Do you value the opportunity to choose or create performance-based assessments to measure student learning? Why do you feel this way?
3. What contributed to your preparation to design (or select) and implement performance assessments?
4. How do you typically ensure a performance assessment can produce appropriate and reliable outcomes?
    a. What is your typical practice of aligning a performance assessment to the content that is taught?
    b. Do you determine if essential resources for a performance assessment are available for all students before implementing a performance assessment? If so, how?
    c. How do you determine that all students clearly understand the communicated procedures of a performance assessment?

d. How do you ensure that students have a range of opportunities for success on a performance assessment?

e. How do you use the results of the performance assessment to directly impact the teaching and learning of your students?

5. Do you feel that the performance-based assessments you implement with your students reflect what students know and are able to do with what they know? If so, how? If you are not sure, what ideas do you have that could assist in planning a professional development opportunity to best support your needs and the needs of other teachers who may not be sure?

APPENDIX H

INFORMED CONSENT AGREEMENT

I agree to participate in one focus group conducted by Lynne Bland, as part of her doctoral research study about the quality of performance assessments. I understand the focus group discussions will last approximately 75 minutes and will focus on understanding the quality of performance assessments implemented with students as one artifact in an alternative portfolio assessment. I will be asked questions about the quality of performance assessments I have implemented with students and my views about the merit (i.e., the inherent value and quality) and worth (i.e., the utility, feasibility, and propriety with their respective teaching contexts), feasibility, and propriety of the selected performance assessments and any other issues I would like to discuss in relation to the topic.

I understand that participation in the focus group is completely voluntary; that I may choose not to answer certain questions, and that I may withdraw and discontinue participation at any time with no negative consequences.

I further understand that my confidentiality will be protected at all times and that a fictitious name will be assigned to me after the focus groups are completed, and that any identifying information/characteristics about me or my family or coworkers will be deleted. The transcripts and the tapes will be assigned a numerical code and or a pseudonym and kept in a password protected computer file. I further understand that the tapes and transcripts created to document my responses will be destroyed after being interpreted and synthesized in writing as the final step of this study.

Focus Group Respondent's Printed Name:_____Signature:_____Date:_____

Focus Group Facilitator_____Signature:_____Date:_____

If you have any further questions please contact me, Lynne Bland, at 804-301-**** or lmband@email.wm.edu.

APPENDIX I

VIRGINIA QUALITY CRITERIA REVIEW TOOL

SAMPLE JUSTIFICATION

Criterion 1 **Standards/Intended Learning Outcomes:**
0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

School	Criterion	USI	USII
1	A	2	2
	B	1	2
	C	1	1
	Quality Rating by Course	4/9 = 44%	5/9 = 56%
Overall Quality Rating	50%		

Summary:
Some of the Standards of Learning and Skills are not identified making it difficult to fully ascertain whether the performance assessments are fully aligned to the content and skills. While the skill of *compare and contrast perspectives* was listed for USII, there was no evidence that students actually engaged with this skill. Limited to partial evidence is provided to show that the performance assessments go beyond simple recall, provide deeper learning competencies or opportunities for collaboration. Students received vague instructions to write a letter in USII which was an attempt at the integration of social studies and language arts. The opportunity for students to engage in life-ready skills appeared to be limited.

APPENDIX J.

Overview Criterion 1

Quality Ratings and Summary of Quality Criteria Review Tool Criterion 1 Data by School

Criterion 1 **Standards/Intended Learning Outcomes** 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence				
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
1	A	2	2	4/2 = 2
	B	1	2	3/2 = 1.5
	C	1	1	2/2 = 1
	Quality Rating by Course	4/3 = 1.3	5/3 = 1.7	
	Overall Quality Rating: 1.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
2	A	0	1	1/2 = .5
	B	3	2	5/2 = 2.5
	C	2	1	3/2 = 1.5
	Quality Rating by Course	5/3 = 1.7	4/3 = 1.3	
	Overall Quality Rating: 1.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
3	A	0	2	2/2 = 1
	B	1	3	4/2 = 2
	C	2	2	4/2 = 2
	Quality Rating by Course	3/3 = 1	7/3 = 2.3	
	Overall Quality Rating: 1.6			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
4	A	0	1	1/2 = .5
	B	1	0	1/2 = .5
	C	1	0	1/2 = .5

	Quality Rating by Course	2/3 = .67	1/3 = .33	
	Overall Quality Rating: 0.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
5	A	2	2	4/2 = 2
	B	3	2	5/2 = 2.5
	C	3	2	5/2 = 2.5
	Quality Rating by Course	8/3 = 2.7	6/3 = 2	
	Overall Quality Rating: 2.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
6	A	3	3	6/2 = 3
	B	2	3	5/2 = 2.5
	C	2	2	4/2 = 2
	Quality Rating by Course	7/3 = 2.3	8/3 = 2.7	
	Overall Quality Rating: 2.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
7	A	3	0	3/2 = 1.5
	B	2	2	4/2 = 2
	C	1	2	3/2 = 1.5
	Quality Rating by Course	6/3 = 2	4/3 = 1.3	
	Overall Quality Rating: 1.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
8	A	0	0	0/2 = 0
	B	1	3	4/2 = 2
	C	1	2	3/2 = 1.5
	Quality Rating by Course	2/3 = .67	5/3 = 1.67	
	Overall Quality Rating: 1.2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
	A	0	0	0/2 = 0

9	B	3	3	6/2 = 3
	C	1	2	3/2 = 1.5
	Quality Rating by Course	4/3 = 1.3	5/3 = 1.7	
	Overall Quality Rating: 1.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
10	A	0	0	0/2 = 0
	B	0	3	3/2 = 1.5
	C	1	2	3/2 = 1.5
	Quality Rating by Course	1/3 = .3	5/3 = 1.7	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
11	A	2	3	5/2 = 2.5
	B	3	3	6/2 = 3
	C	3	3	6/2 = 3
	Quality Rating by Course	8/3 = 2.7	9/3 = 3	
	Overall Quality Rating: 2.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
12	A	3	0	3/2 = 1.5
	B	3	3	6/2 = 3
	C	3	2	5/2 = 2.5
	Quality Rating by Course	9/3 = 3	5/3 = 1.7	
	Overall Quality Rating: 2.4			

Note. United States History to 1865 (USI) and United States History 1865 to the present (USII)

Appendix K

Overview Criterion 2

Quality Rating and Summary of Quality Criteria Review Tool Criterion 2 Data by School

Criterion 2 **Authenticity** 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence				
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
1	A	1	1	2/2 = 1
	Quality Rating by Course	1/1 = 1	1/1 = 1	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
2	A	3	3	6/2 = 3
	Quality Rating by Course	3/1 = 3	3/1 = 3	
	Overall Quality Rating: 3			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
3	A	1	3	4/2 = 2
	Quality Rating by Course	1/1 = 1	3/1 = 3	
	Overall Quality Rating: 2			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
4	A	2	1	3/2 = 1.5
	Quality Rating by Course	2/1 = 2	1/1 = 1	
	Overall Quality Rating: 1.5			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
5	A	3	3	6/2 = 3
	Quality Rating by Course	3/1 = 3	3/1 = 3	
	Overall Quality Rating: 3			

School	Criterion	USI	USII	Quality Ratings
6	A	3	3	6/2 = 3
	Quality Rating by Course	3/1 = 3	3/1 = 3	
	Overall Quality Rating: 3			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
7	A	2	3	5/2 = 2.5
	Quality Rating by Course	2/1 = 2	3/1 = 3	
	Overall Quality Rating: 2.5			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
8	A	1	3	4/2 = 2
	Quality Rating by Course	1/1 = 1	3/1 = 3	
	Overall Quality Rating: 2			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
9	A	3	2	5/2 = 2.5
	Quality Rating by Course	3/1 = 3	2/1 = 2	
	Overall Quality Rating: 2.5			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
10	A	1	2	3/2 = 1.5
	Quality Rating by Course	1/1 = 1	2/1 = 2	
	Overall Quality Rating: 1.5			
School	**Criterion**	**USI**	**USII**	**Quality Ratings**
11	A	3	3	6/2 = 3
	Quality Rating by Course	3/1 = 3	3/1 = 3	
	Overall Quality Rating: 3			

School	Criterion	USI	USII	Quality Ratings
12	A	3	3	6/2 = 3
	Quality Rating by Course	3/1 = 3	3/1 = 3	
	Overall Quality Rating: 3			

Note. United States History to 1865 (USI) and United States History 1865 to the present (USII)

APPENDIX L

Overview Criterion 3

Quality Rating and Summary of Quality Criteria Review Tool Criterion 3 Data by School

Criterion 3: **Language Use for Expressing Reasoning** 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence				
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
1	A	2	3	5/2 = 2.5
	B	3	1	4/2 = 2
	Quality Rating by Course	5/2 = 2.5	4/2 = 2	
	Overall Quality Rating: 2.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
2	A	3	3	6/2 = 3
	B	3	3	6/2 = 3
	Quality Rating by Course	6/2 = 3	6/2 = 3	
	Overall Quality Rating: 3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
3	A	3	3	6/2 = 3
	B	3	2	5/2 = 2.5
	Quality Rating by Course	6/2 = 3	5/2 = 2.5	
	Overall Quality Rating: 2.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
4	A	1	1	2/2 = 1
	B	1	1	2/2 = 1

	Quality Rating by Course	2/2 = 1	2/2 = 1	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
5	A	3	1	4/2 = 2
	B	3	1	4/2 = 2
	Quality Rating by Course	6/2 = 3	2/2 = 1	
	Overall Quality Rating: 2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
6	A	3	3	6/2 = 3
	B	2	1	3/2 = 1.5
	Quality Rating by Course	5/2 = 2.5	4/2 = 2	
	Overall Quality Rating: 2.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
7	A	2	3	5/2 = 2.5
	B	2	3	5/2 = 2.5
	Quality Rating by Course	4/2 = 2	6/2 = 3	
	Overall Quality Rating: 2.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
8	A	1	2	3/2 = 1.5
	B	2	1	3/2 = 1.5
	Quality Rating by Course	3/2 = 1.5	3/2 = 1.5	
	Overall Quality Rating: 1.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**

9	A	2	3	5/2 = 2.5
	B	1	1	2/2 = 1
	Quality Rating by Course	3/2 = 1.5	4/2 = 2	
	Overall Quality Rating: 1.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
10	A	0	2	2/2 = 1
	B	1	2	3/2 = 1.5
	Quality Rating by Course	1/2 = .5	4/2 = 2	
	Overall Quality Rating: 1.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
11	A	3	3	6/2 = 3
	B	3	3	6/2 = 3
	Quality Rating by Course	6/2 = 3	6/2 = 3	
	Overall Quality Rating: 3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
12	A	3	3	6/2 = 3
	B	1	3	4/2 = 2
	Quality Rating by Course	4/2 = 2	6/2 =3	
	Overall Quality Rating: 2.5			

Note. United States History to 1865 (USI) and United States History 1865 to the present (USII)

APPENDIX M

Overview Criterion 4

Quality Rating and Summary of QCRT Criterion 4 Data by School

Criterion 4: **Success Criteria for Students** 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence				
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
1	A	1	1	2/2 = 1
	B	0	0	0
	C	0	0	0
	Quality Rating by Course	1/ = .3	1/3 = .3	
	Overall Quality Rating: 0.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
2	A	1	0	1/2 = .5
	B	3	0	3/2 = 1.5
	C	2	0	2/2 = 1
	Quality Rating by Course	6/3 = 2	0	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
3	A	1	0	1\2 = .5
	B	1	0	1\2 = .5
	C	0	0	0
	Quality Rating by Course	2/3 = .7	0	
	Overall Quality Rating: 0.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**

4	A	1	1	2/2 = 1
	B	1	1	2/2 = 1
	C	1	0	1/2 = .5
	Quality Rating by Course	3/3 = 1	2/3 = .7	
	Overall Quality Rating: 0.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
5	A	1	0	1/2 = .5
	B	1	0	1/2 = .5
	C	1	0	1/2 = .5
	Quality Rating by Course	3/3 = 1	0/3 = 0	
	Overall Quality Rating: 0.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
6	A	3	2	5/2 = 2.5
	B	2	3	5/2 = 2.5
	C	1	1	2/2 = 1
	Quality Rating by Course	6/3 = 2	6/3 = 2	
	Overall Quality Rating: 2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
7	A	2	2	4/2 = 2
	B	2	2	4/2 = 2
	C	2	3	5/2 = 2.5
	Quality Rating by Course	6/3 = 2	7/3 = 2.3	
	Overall Quality Rating: 2.2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**

8	A	2	1	3/2 = 1.5
	B	1	2	3/2 = 1.5
	C	0	0	0
	Quality Rating by Course	3/3 = 1	3/3= 1	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
9	A	1	1	2/2 = 1
	B	1	1	2/2 = 1
	C	0	0	0
	Quality Rating by Course	2/3 = .7	2/3 = .7	
	Overall Quality Rating: 0.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
10	A	0	0	0
	B	0	0	0
	C	0	0	0
	Quality Rating by Course	0	0	
	Overall Quality Rating: 0			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
11	A	0	3	3/2 = 1.5
	B	0	2	2/2 = 1
	C	0	2	2/2 = 1
	Quality Rating by Course	0	7/3 = 2.3	
	Overall Quality Rating: 1.2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**

12	A	0	1	1/2 = .5
	B	0	1	1/2 =.5
	C	0	0	0
	Quality Rating by Course	0	2/3 = .7	
	Overall Quality Rating: .3			

Note: United States History to 1865 (USI) and United States History 1865 to the present (USII)

APPENDIX N

Overview Criterion 5

Quality Rating, and Summary of Quality Criteria Review Tool Criterion 5 Data by School

Criterion 5: **Student Directions, Prompt, and Resources/Materials** 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.				
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
1	A	0	0	0
	B	2	0	2/2 = 1
	C	3	0	3/2 = 1.5
	Quality Rating by Course	5/3 = 1.7	0	
	Overall Quality Rating: 0.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
2	A	2	2	4/2 = 2
	B	2	2	4/2 = 2
	C	3	3	6/2 = 3
	Quality Rating by Course	7/3 = 2.3	7/3 = 2.3	
	Overall Quality Rating: 2.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
3	A	1	2	3/2 = 1.5
	B	1	2	3/2 = 1.5
	C	3	3	6/2 = 3
	Quality Rating by Course	5/3 = 1.7	7/3 = 2.3	
	Overall Quality Rating: 2			
School	**Criterion**	**USI**	**USII**	**sub-criterion Quality Ratings**
4	A	1	1	2/2 = 1

	B	2	1	3/2 = 1.5
	C	3	0	3/2 = 1.5
	Quality Rating by Course	6/3 = 2	2/3 = .7	
	Overall Quality Rating: 1.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
5	A	0	2	2/2 = 1
	B	3	2	5/2 = 2.5
	C	3	3	6/2 = 3
	Quality Rating by Course	6/3 = 2	7/3 = 2.3	
	Overall Quality Rating: 2.2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
6	A	2	2	4/2 = 2
	B	2	3	5/2 = 2.5
	C	1	3	4/2 = 2
	Quality Rating by Course	5/3 = 1.7	8/3 = 2.7	
	Overall Quality Rating: 2.2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
7	A	2	2	4/2 = 2
	B	2	2	4/2 = 2
	C	3	3	6/2 = 3
	Quality Rating by Course	7/3 = 2.3	7/3 = 2.3	
	Overall Quality Rating: 2.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
8	A	2	2	4/2 = 2
	B	1	3	4/2 = 2

	C	0	3	3/2 = 1.5
	Quality Rating by Course	3/3 = 1	8/3 = 2.7	
	Overall Quality Rating: 1.9			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
9	A	1	0	1/2 = .5
	B	2	2	4/2 = 2
	C	0	3	3/2 = 1.5
	Quality Rating by Course	3/3 = 1	5/3 = 1.7	
	Overall Quality Rating: 1.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
10	A	0	0	0
	B	1	1	2/2 = 1
	C	0	3	3/2 = 1.5
	Quality Rating by Course	1/3 = .3	4/3 = 1.3	
	Overall Quality Rating: 0.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
11	A	2	2	4/2 = 2
	B	3	3	6/2 = 3
	C	3	3	6/2 = 3
	Quality Rating by Course	8/3 = 2.7	8/3 = 2.7	
	Overall Quality Rating: 2.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
12	A	2	1	3/2 = 1.5
	B	1	1	2/2 = 1
	C	3	3	6/2 = 3
	Quality Rating by Course	6/3 = 2	5/3 = 1.7	

	Overall Quality Rating: 1.8

Note: United States History to 1865 (USI) and United States History 1865 to the present (USII)

APPENDIX O

Overview Criterion 6

Quality Rating and Summary of Quality Criteria Review Tool Criterion 6 Data by School

Criterion 6: **Accessibility**				
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
1	A	0	0	0
	B	2	0	2/2 = 1
	Quality Rating by Course	2/2 = 1	0	
	Overall Quality Rating: 0.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
2	A	0	0	0
	B	0	1	1/2 = .5
	Quality Rating by Course	0	1/2 = .5	
	Overall Quality Rating: 0.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
3	A	0	1	1/2 = .5
	B	1	2	3/2 - 1.5
	Quality Rating by Course	1/2 = .5	3/2 = 1.5	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
4	A	0	0	0
	B	0	0	0
	Quality Rating by Course	0	0	

	Overall Quality Rating: 0			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
5	A	0	0	0
	B	2	1	3/2 = 1.5
	Quality Rating by Course	2/2 = 1	1/3 = .3	
	Overall Quality Rating:.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
6	A	1	0	1/2 =.5
	B	2	0	2/2 = 1
	Quality Rating by Course	3/2 = 1.5	0	
	Overall Quality Rating:.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
7	A	2	1	3/2 = 1.5
	B	1	1	2/2 = 1
	Quality Rating by Course	3/2 = 1.5	2/2 = 1	
	Overall Quality Rating: 1.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
8	A	1	2	3/2 = 1.5
	B	0	1	1/2 =.5
	Quality Rating by Course	1/2 = .5	3/2 = 1.5	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
9	A	2	0	2/2 = 1

	B	1	2	3/2 = 1.5
	Quality Rating by Course	3/2 = 1.5	2/2 = 1	
	Overall Quality Rating: 1.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
10	A	0	1	1/2 =.5
	B	0	3	3/2 = 1.5
	Quality Rating by Course	0	4/2 = 2	
	Overall Quality Rating: 1			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
11	A	0	3	3/2 = 1.5
	B	3	3	6/2 = 3
	Quality Rating by Course	3/2 = 1.5	6/2 = 3	
	Overall Quality Rating: 2.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
12	A	2	1	3/2 = 1.5
	B	0	2	2/2 = 1
	Quality Rating by Course	2/2 = 1	3/2 = 1.5	
	Overall Quality Rating: 1.3			

Note. Scale: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence; United States History to 1865 (USI) and United States History 1865 to the present (USII).

APPENDIX P

Overview Criterion 7

Quality Rating and Summary of Criterion 7 Data by School

Criterion 7: **Feasibility**				
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
1	A	2	0	2/2 = 1
	B	0	0	0
	C	0	0	0
	Quality Rating by Course	2/3 = .7	0	
	Overall Quality Rating: 0.3			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
2	A	1	1	2/2 = 1
	B	0	1	1/2 =.5
	C	0	0	0
	Quality Rating by Course	1/3 = .3	2/3 = .7	
	Overall Quality Rating: 0.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
3	A	1	2	3/2 = 1.5
	B	1	1	2/2 = 1
	C	0	0	0
	Quality Rating by Course	2/3 = .7	3/3 = 1	
	Overall Quality Rating: 0.8			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
4	A	2	1	3/2 = 1.5
	B	0	0	0

	C	0	0	0
	Quality Rating by Course	2/3 = .7	1/3 = .3	
	Overall Quality Rating: 0.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
5	A	3	1	4/2 = 2
	B	3	0	3/2 - 1.5
	C	3	0	3/2 = 1.5
	Quality Rating by Course	9/3 = 3	1/3 = .3	
	Overall Quality Rating: 1.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
6	A	2	2	4/2 = 2
	B	0	0	0
	C	0	0	0
	Quality Rating by Course	2/3 = .7	2/3 = .7	
	Overall Quality Rating: 0.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
7	A	3	3	6/2 = 3
	B	3	3	6/2 = 3
	C	0	3	3/2 = 1.5
	Quality Rating by Course	6/3 = 2	9/3 = 3	
	Overall Quality Rating: 2.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
8	A	2	2	4/2 = 2
	B	3	0	3/2 = 1.5
	C	0	0	0

	Overall Quality Rating	5/3 = 1.7	2/3 = .7	
	Overall Quality Rating: 1.2			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
9	A	2	1	3/2 = 1.5
	B	0	0	0
	C	0	0	0
	Quality Rating by Course	2/3 = .7	1/3 = .3	
	Overall Quality Rating: 0.5			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
10	A	1	1	2/2 = 1
	B	1	0	1/2 = .5
	C	0	1	1/2 = .5
	Quality Rating by Course	2/3 = .7	2/3 = .7	
	Overall Quality Rating: 0.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
11	A	1	3	4/2 = 2
	B	0	0	0
	C	0	0	0
	Quality Rating by Course	1/3 =.3	3/3 = 1	
	Overall Quality Rating:.7			
School	**Criterion**	**USI**	**USII**	**Sub-criterion Quality Ratings**
12	A	0	2	2/2 = 1
	B	2	0	2/2 = 1
	C	2	0	2/2 = 1
	Quality Rating by Course	4/3 = 1.3	2/3 = .7	

	Overall Quality Rating: 1

Note. 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence;

United States History to 1865 (USI) and United States History 1865 to the present (USII).

APPENDIX Q

2015 History and Social Science Standards of Learning Skills Progression by Grade or Course

2015 History and Social Science Standards of Learning Skills Progression by Grade or Course

Conceptual Understanding: The student is first introduced to the skill and applies the skill appropriately in varied situations.
Scaffolding the Understanding: The student continues to develop the skill and integrates the skill appropriately across new concepts.
Analyzing the Understanding: The student is knowledgeable about the skill from previous instruction, applies the skill to new concepts, and analyzes outcomes

Social Science Skill	Description of Skill	Grade/Course												
		K	1	2	3	VS	USI	USII	CE	WG	WHI	WHII	VUS	GOVT
1a Using information sources	View artifacts, primary and secondary sources													
	Use artifacts, primary and secondary sources													
	Identify artifacts, primary and secondary sources													
	Analyze and interpret artifacts, primary and secondary sources													
	Synthesize evidence from primary and secondary sources													
1b Applying geographic skills	Use basic map skills													
	Use geographic information													
	Analyze the impact of geographic features													
	Use geographic information to determine patterns and trends													
1c Organizing information	Gather and classify information, sequence events, and separate fact from fiction													
	Use and create charts, graphs, diagrams, and pictures to determine characteristics of people, places or events													
	Interpret charts, graphs, diagrams, and pictures to determine characteristics of people, places or events													
1d Questioning and using critical thinking skills	Ask appropriate questions to solve a problem													
	Summarize points and evidence to answer a question													
	Recognize points of view and historical perspective													
	Use evidence to draw conclusions and make generalizations													
	Evaluate sources for accuracy, credibility, bias, and propaganda													
	Construct arguments using evidence from multiple sources													
1e Comparing and Contrasting	Compare and contrast people, places, or events													
	Compare and contrast ideas and perspectives													
	Compare and contrast historical, cultural, and political perspectives													

2015 History and Social Science Standards of Learning Skills
Progression by Grade or Course

Social Science Skill	Description of Skill	Grade/Course												
		K	1	2	3	VS	USI	USII	CE	WG	WHI	WHII	VUS	GOVT
1f Determining cause-and-effect	Recognize direct cause-and-effect relationships													
	Determine relationships with many causes or effects													
	Explain indirect cause-and-effect relationships													
1g Making connections	Make connections between past and present													
	Explain connections across time and place													
	Analyze multiple connections across time and place													
1h Making economic decisions	Use a decision-making model to make informed decisions													
	Use a decision-making model to identify costs and benefits of a specific choice made													
	Use a decision-making model to analyze and explain the incentives and consequences of a specific choice made													
1i Exercising civic responsibility	Practice good citizenship skills while collaborating, compromising, and participating in classroom activities													
	Demonstrate respect for rules and laws													
	Identify the rights and responsibilities of citizenship													
	Identify ethical use of material or intellectual property													
1j Demonstrating comprehension	Develop fluency in content vocabulary, and comprehension of verbal, written, and visual sources													
	Defend positions using content vocabulary													
	Access a variety of media, including online resources													
	Investigate and research to develop products orally and in writing													

APPENDIX R

Middle School Common Rubric Core Expectations

	4	3	2	1	Not Observed
Core Expectations (.1a and .1c)					
Accuracy of Content Synthesizing information sources Explaining Evidence	• Identified, analyzed, and interpreted information sources to demonstrate in-depth understanding of content • Integrated evidence from a variety of information sources to determine characteristics of people, places, events, or concepts • Used information to consistently develop, support, or sharpen the explanation or statement	• Analyzed and interpreted information sources to understand specific content • Gathered and classified information to sequence events and separate fact from fiction • Used information to develop and support an explanation or statement	• Used information sources to understand concepts, people, places, or events • Classified information, sequenced events, and separated fact from fiction • Used information to support an explanation	• Used information sources to understand content • Separated fact from fiction • Identified information to support an explanation	

APPENDIX S

Middle School Common Rubric Task Specific Concepts and Skills

Task Specific Concepts and Skills					
	4	**3**	**2**	**1**	**Not Observed**
Geographic Patterns and Trends (.1b)	Used geographic information to analyze the impact of geographic features on a pattern or trend.	Used basic map skills and geographic information to identify a pattern or trend in data	Used basic map skills to identify data	Used basic map skills	
Evaluating Sources (.1d)	Used evidence to draw conclusions and make generalizations about points of view and historical perspective	Used evidence to summarize points of view or historical perspective	Used evidence identify points of view or historical perspective	Answered questions about points of view or historical perspective	
Explanation or Statement (.1d)	Responded to the task with a decisive explanation or statement beyond conventional conclusions	Responded to the task with a reasonable explanation or statement	Responded to the task with a partially developed explanation or statement	Attempted to present a central explanation or statement	
Differing Perspectives (.1e)	Compared and contrasted ideas about historical, cultural and political perspectives in history	Compared and contrasted concepts, people, places, or events	Explained concepts, people, places, or events	Identified concepts, people, places, or events	

	4	**3**	**2**	**1**	**Not Observed**
Determine causes or effects (.1f)	Determined and explained relationships with many causes or effects	Explained direct cause-and-effect relationships	Identified direct cause-and-effect relationships	Identified a cause-and-effect relationship	
Connections across time (.1g)	Explained connections across time and place	Made connections between past and present events	Made connections between past events	Identified past and present events	
Making decisions (.1h)	Used a decision-making model identify the costs and benefits of a specific choice made	Identified the costs and benefits of a specific choice	Identified the costs or benefits of a specific choice made	Identified that a specific choice was made	
Citizenship (.1i)	Used authentic, valid sources and gave credit when using outside ideas, opinions, or theories	Used sources and gave credit when using outside ideas, opinions, or theories.	Used sources and gave credit incorrectly when using another person's ideas, opinions, or theories	Used sources	
Developing Research Questions (.1j)	Identified a question and made a connection between the question and existing information or ideas about a topic	Identified a question and stated existing ideas or information about a topic	Restated existing ideas or information about a topic	Made up ideas or information about a topic	
Selecting Sources (.1j)	Selected relevant sources by accessing a variety of media, including online resources	Selected sources from a variety of media	Selected sources that represent two different types of media	Selected sources	

Vita
Lynne Mardigian Bland

Birthdate:	September 29, 1963	
Birthplace:	Richmond, Virginia	
Education:	2011 - 2022	College of William and Mary Williamsburg, VA Doctor of Education in Curriculum Leadership
	2005 - 2007	University of Virginia Charlottesville, Virginia Master of Education in Reading Education
	1981 - 1985	Virginia Commonwealth University Richmond, Virginia Bachelor of Science in Elementary Education
Professional Experience	2012 – Present	Social Studies Curriculum Specialist Chesterfield County Public Schools Chesterfield, VA
	2005 – 2012	Social Studies Teacher Consultant Chesterfield County Public Schools Chesterfield, VA
	2011 – Present	Adjunct Instructor University of Richmond University of Richmond, VA
	2009 – 2009	Adjunct Instructor Virginia Commonwealth University Richmond, VA
	2002 – 2005	Middle School Social Studies Teacher Chesterfield County Public Schools Chesterfield, VA
	1993 - 2002 1986 - 1991	Middle School Social Studies Teacher Henrico County Public Schools Henrico, VA
	1991 – 1993	Case Manager Southern Highlands Community Mental Health Center Mullens, WV