Spring 3-10-2021

An Analysis of the Higher Order Thinking Requirements of PARCC Practice Assessments in Grades 10

Heather Dorrian
heather.dorrian@student.shu.edu

An Analysis of the Higher Order Thinking Requirements of PARCC Practice Assessments in Grades 10

Heather Dorrian

Dissertation Committee
Christopher Tienken Ed.D
Leslie A. Solis-Stovall Ed.D
Michael Kuchar Ed.D

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Education

Department of Educational Leadership, Management, and Policy

Seton Hall University
2021

SETON HALL UNIVERSITY

1856

COLLEGE OF EDUCATION & HUMAN SERVICES
DEPARTMENT OF EDUCATION LEADERSHIP MANAGEMENT & POLICY

APPROVAL FOR SUCCESSFUL DEFENSE

Heather Dorrian has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Spring** Semester.

DISSERTATION COMMITTEE
(please sign and date)

Mentor	**Date**
Christopher Tienken Ed.D	5/11/2021
Committee Member	**Date**
Leslie A. Solis-Stovall Ed.D	5/11/2021
Committee Member	**Date**
Michael Kuchar Ed.D	5/11/2021
Committee Member	**Date**

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation.

Abstract

This mixed methods study aimed to categorize and analyze the frequencies and percentages of complex thinking in the PARCC practices assessments in English Language Arts grade 10 and Geometry. The Hess' Cognitive Rigor Matrix was used for the first part of the study to code each of the PARCC assessment questions in Language Arts grade 10 and Geometry based on pre-existing codes. Deductive category application was utilized to connect the language from Hess' Cognitive Rigor Matrix to the language of the questions in the tests. To ensure reliability we utilized the double-rater read behind method as in other similar studies. In the second part of the study, a quantitative methods approach was implemented to determine the frequencies. Moreover, descriptive statistics was then utilized to describe the differences and similarities of complex thinking that exist in the language of the PARCC practice assessment. In response to the research questions, the data analyzed revealed the following trends from the Language Arts in grade 10 and Geometry PARCC Practice Tests:

1. The questions in the Language Arts PARCC tests in grade 10 were rated at an overall higher percentage for lower-level questions.

2. The questions in the Geometry PARCC tests were rated at an overall higher percentage for lower-level questions.

3. No questions were placed at the most cognitive complex level. This study suggests that more opportunities for developing complex thinking, which is essential to 21st century learning, is implemented through standardized assessments.

Keywords: *Higher Order Thinking. Critical Thinking. Complex Thinking. 21st century skills. Partnership for Assessment of Readiness for College and Career (PARCC). Common Core Standards.*

Dedication and Acknowledgements

This dissertation is culminative experience of my education and personal endeavors to continue to strive to learn and develop. Much like any journey, I would not have made it to this point without help from loved ones and friends along the way.

I am internally indebted to my friend, Dr. Leslie Solis who dragged me up this hill kicking and screaming. Leslie is the type of person who truly embodies the educator's spirit. With her guidance and support, she helped me to set my mind upon finalizing this dissertation. I am so thankful for her and her eagerness to support me and her unyielding authenticity. Despite losing her father to Covid-19 in 2020 and raising a family of three, she still took time out to coach and encourage me throughout this process. It is because of her support that I dedicate this dissertation to her as small way of thanks.

I especially want to thank my family and friends for their ongoing support and love throughout this process. Thank you to a very special person in my life, Lisa Gentile, for your unfailing love and support throughout this journey. Thank you for always making me believe in the choices I make and for encouraging me along the way. Thank you so much to my mom, Rosemary Dorrian, for always being my champion and friend. You are a wonderful and caring mother. Thanks also to Jodi Curiotto, Luisa Surma and Keri English for their support and friendship.

To the family members and loved ones that I lost in the process of this dissertation, you will forever be in my heart. Especially Lynne and Bob Ely you were such a big part of my life and were taken from us too soon. Without your love and presence in my life, I would not be the person that I am today. To my friend Eileen Antonison, while our friendship was all too brief, your infectious spirit, vigor and passion for education as well as helping others will always remain behind for those who encountered you. I will always carry our friendship with me.

I especially want to thank my committee who I am incredibly grateful for. To my mentor, Dr. Tienken, this work would not have been possible without your guidance. I remember the first days of class being so struck by the vastly different perspective and lens you shared in regards to policy and education. Your sharp intellect, criticism and way in which you view policy and curriculum has shaped the ways in which I will forever view education. Your passion for education and changing outcomes is truly inspiring.

My earnest thanks to Dr. Armand Lamberti, who was another pivotal figure on my journey to completing my doctorate. Throughout the program, not only was Armand a friend but he was the needle in my side that made me rethink things my conclusions with an ever-increasing critical eye. His perspectives helped to sharpen my thought process, as well as helped me gain a friend along my journey. I am very grateful for your involvement in the coding process and your friendship throughout the program.

I am also grateful for Dr. Kuchar and his counsel throughout the course of my research which led me to the successful completion of my dissertation.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

Background

In May 1996, the New Jersey State Board of Education adopted the Core Curriculum Content Standards (CCCS) that enumerated what all New Jersey students should know and be able to do by the end of the fourth and eighth grades, and upon completion of a New Jersey public school education. The CCCS, which are revised every five years, also define New Jersey's high school graduation requirements and are the basis for assessing the academic achievement of students at grades 3 through 12 (State of New Jersey Department of Education, 2016). The CCCS informed the development of three subsequent statewide assessments: the Elementary School Proficiency Assessment (ESPA) that was administered from 1997-2002; the Grade Eight Proficiency Assessment (GEPA), which replaced the EWT in 1998 and was administered through 2007-2008; and the High School Proficiency Assessment (HSPA), that replaced the HSPT11 as the state's graduation test for all students who entered the eleventh grade in the fall of 2001.

With the enactment of the No Child Left Behind Act of 2001 (NCLB), New Jersey's statewide assessment system underwent further change. This federal legislation required that each state administer annual standards-based assessments to students in grades 3 through 8, and at least once in high school. The federal expectation was that each state would provide tests that were grounded in rigorous state content standards and that would assess student achievement in language arts literacy, mathematics and, at three benchmark grade levels, science.

In response to *NCLB* requirements and New Jersey's own expectations that students would be reading at grade level by the end of third grade, New Jersey revised its elementary assessment to include a third-grade assessment program. The *New Jersey Assessment of Skills and Knowledge (*NJ ASK*)* was field-tested in May 2003, became fully operational the following year. With the implementation of NJ ASK 3 in 2003, the ESPA became the NJ ASK4. The state's elementary science assessment was first administered to New Jersey's fourth graders in spring 2004, becoming operational the following year. NJ ASK was further expanded in 2006 to include grades 5 through 7. New Jersey's assessment system then included NJ ASK 3-8, HSPA, the Alternate Proficiency Assessment (APA) for students with severe cognitive disabilities, and end-of-course high school competency assessments in biology and algebra.

In June 2010, the New Jersey State Board of Education adopted the Common Core State Standards (CCSS) in mathematics and English language arts/literacy. In 2011, The NJ Department of Education submitted a waiver application to the US Department of Education for relief from certain provisions of No Child Left Behind *(NCLB).* The comprehensive waiver allowed the Department to develop a new accountability system to replace the provisions of *NCLB*, centered on providing support and intervention to the state's lowest-performing schools and those with the largest in-school gaps between subgroups of students.

In preparation for the new accountability system, the state joined the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium in the spring of 2010. New Jersey became a Governing State in the spring of 2011 and actively helped shape PARCC's proposal for a common, next-generation assessment system.

The Partnership for Assessment of Readiness for College and Careers (PARCC) is a consortium of states that collaboratively developed a common set of assessments to measure student achievement of the Common Core State Standards and preparedness for college and careers. In 2014-15, the PARCC electronic assessments replaced the previous statewide assessments -- the NJ ASK in grades 3-8 and HSPA in high school. New Jersey had been transitioning the NJ ASK to measure higher-level skills for over three years to provide local districts and schools the time necessary to shift practices and prepare students and educators for PARCC.

In May, 2016, the NJ State Board of Education adopted the revised mathematics and English language arts standards and changed the name of all nine areas of New Jersey's standards to the NJ Student Learning Standards. The PARCC assessments are aligned to high-level thinking skills and were created to measure students' ability to apply their knowledge of concepts rather than repeat memorized facts. The PARCC assessments require students to solve problems using mathematical reasoning and to be able to model mathematical principles. In English language arts (ELA), students are required to closely read multiple passages and to write essay responses in literary analysis, research tasks and narrative tasks. The assessments also provide teachers and parents with information on student progress to inform instruction and provide targeted student support.

Background

The Race to the Top assessments aligned with the Common Core State Standards; which include the PARCC exam, have been extensively examined technically but less has been done on analyzing the type of critical analytical thinking (CAT) found within these assessments: (Brown et al., 2014).

Therefore, the assessment that measures these standards should reflect and demand cognitive rigor. This summative assessment serves to illustrate a student's progress in regards to their understanding and development of these standards and skills. Little empirical evidence exists that describe how the language found in the English Language Arts and Mathematics sections of the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in High School compare with the language that promotes higher-order thinking found in research literature.

The PARCC test claims to emphasize a marked increase in critical analytical thinking and higher order thinking skills. Generally, higher order thinking skills are thinking processes that require more than simple recall from learners. The Common Core State Standards, and the Next Generation Science Standards, emphasizes the need to assess CAT and to be able to develop a transfer of skills across subject matters. For Grade 11 High School English Language Arts, for instance, the PARCC consortium claims that, "The PARCC complexity framework reflects the importance of text complexity as it relates to the CCSS, (Common Core Curriculum Standards), which indicates that 50 percent of an item's complexity is linked to the complexity of the text(s) used as the stimulus for that item To this end, PARCC has developed a clear and consistent model to define text complexity and has determined to use three text complexity levels: readily accessible, moderately complex, or very complex" (PARCC, ETS, PEARSON, Assessment SIG Business Meeting , 2014). Three main sources of item complexity on this test include: (a) Command of textual evidence, (b) Response mode, and (c) processing demand; processing Complexity is combined with the fourth source, Text Complexity, to produce an overall cognitive complexity score.

However, current assessments often do not develop appropriate frameworks to establish a valid measure that assesses higher order thinking skills. Moreover, measurement strategies such as, subjective rubrics and item constructions, do not often provide the appropriate tools to reflect a learner's true progress (Development Process, 2018). A principled approach to assessment design is critical to ensure accuracy, as well as cognitive rigor that mirrors a student's ability to transfer skills.

Cognitive Rigor

Cognitive rigor is marked and measured by the depth and extent students are challenged and engaged to demonstrate and communicate their knowledge and thinking. It also marks and measures the depth and complexity of student learning experiences (Francis, 2017).

In fact, current methodologies for judging the alignment between standards and assessment routinely incorporate judgments about the quality of the match in both content and cognitive rigor. Study after study has shown that current tests focus on lower levels of knowledge and application at the expense of those addressing deeper learning and high levels of cognitive demand. Yuan and Li (2012), for example, analyzed the intellectual demands of released items and tasks from 17 leading states that were reputed to have the most cognitively demanding tests in the nation. Using Norman Webb's (2007) four-point depth-of-knowledge (DOK) framework to examine cognitive complexity, these researchers found the vast preponderance of selected response items to be at levels 1 and 2 (Herman et al., 2014). This illustrates that the demand for cognitive rigor does not always translate to cognitive rigor or higher order thinking within students.

Higher Order Thinking

Higher order thinking skills are thinking processes that require more than simple recall of facts (deBono, 1983). In order to use higher order thinking skills a student must be able to use metacognitive strategies. Thinking skills such as problem solving, decision making, and conceptualizing, evaluating, synthesizing, and creative thinking require students to evaluate, plan, and monitor their thinking continuously (Lipman, 1991). Higher order thinking (HOT), takes thinking to higher levels than restating the facts. HOT requires that students do something with the facts. For instance, they might understand them, infer from them, connect them to other facts and concepts, categorize them, manipulate them, put them together in new or novel ways, and apply them as they seek new solutions to new problems.

Frameworks of Higher Order Thinking

Various models exist that attempt to explain and measure the types of tasks and problems that should be given to further cultivate higher order thinking. The two most notable frameworks are (1) Bloom's Revised Taxonomy and (2) Webb's Depth of Knowledge. Then in 2005, Karen Hess and her colleagues developed a model that superimposed these two frameworks.

In 1956, Bloom's Taxonomy was created by Benjamin Bloom, published a classification of learning outcomes that are supposed to encourage higher level thinking within learners. Most notably, Bloom's taxonomy differentiates between cognitive skill levels that is supposed to be conducive to creating a greater variety of tasks and assessments that facilitate higher order thinking within various contexts (Adams, 2015). Bloom's taxonomy currently contains six categories of cognitive skills ranging from lower-order skills that require less cognitive processing to higher-order skills that require deeper learning and a greater degree of cognitive processing. The differentiation into categories of higher-order and lower-order skills arose later. (Adams, 2015).

While Bloom's Taxonomy still is prevalent throughout the United States both in assessment planning, as well as classroom instructional practice its model was not intended for its current uses. Moreover, it is limited in truly elevating higher order thinking in its current application. Roland Case, from The Critical Thinking Consortium, emphasizes that "assessing students' ability to complete the "higher order" tasks does not logically imply that students have mastered the "lower order task," (Parker, 2015, p. 75).

There is an erroneous assumption that "lower order thinking tasks" are less demanding than "higher order thinking" tasks. Moreover, learning a topic is not dependent on levels of complexity that scaffold students towards what is labeled "higher level thinking." Instead, Case argues that invited students to make reasoned judgments is a much more productive form of framing learning tasks outside of using jargoned verbs in attempts to move students towards a higher level of critical thinking. Even the revised version of Bloom has some shortcomings: sometimes verbs and processes can seem similar in differing levels; and thinking process, even at higher levels, do not necessarily translate to deeper understanding of content.

Since Bloom's Taxonomy other frameworks have evolved to help further instructional decisions, cultivate critical thinking and cognitive demand in instructional practice. One notable framework is Webb's Depth of Knowledge. In 1997, Norman Webb designed his model as a means of increasing the cognitive complexity and demand of standardized assessments. Webb argued that higher order thinking correlates to the kind of knowledge and type of thinking that needs to be demonstrated in order to answer a question, address a problem, or accomplish a task. While Webb draws on Bloom's taxonomy in his development of learning, depth of knowledge is an entirely different means of measuring and monitoring rigorous learning. It correlates more to how extensively students are able to express and share their knowledge and thinking. Depth of knowledge deals with the setting, scenario, or situation in which thinking is demonstrated and it addresses context rather than cognition.

This framework helped to create another important perspective of cognitive complexity. Webb's work highlights that both the content assessed in a test item and the intended cognitive demand, or the depth to which we expect students to demonstrate understanding of that content, is instrumental in aligning appropriate outcomes in assessments. Meaning, both the complexity and content are needed to determine the depth of knowledge levels (Hess et al., 2009).

While Webb's Depth of Knowledge is nominative rather than a hierarchical classification like that of Bloom's, Webb's explains that depth of knowledge "can vary on a number of dimensions, including level of cognitive complexity of information students should be expected to know, how well they should be able to transfer this knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas". Depth of knowledge is actually about context. It involves the different scenarios, setting, and situations that knowledge can be transferred and used.

Unfortunately, depth of knowledge has been misinterpreted as similar to higher order thinking. Higher order thinking correlates to the kind of knowledge and type of thinking that needs to be demonstrated in order to answer a question, address a problem, or accomplish a task. However, Depth of knowledge is an entirely different means of measuring and monitoring rigorous teaching and learning. It correlates more with how extensively students are able to express and share their knowledge and thinking. Depth of knowledge establishes the setting, scenario, or situation in which learning is demonstrated and communicated. Webb's Depth-of-

Knowledge Model, helps to establish the criteria to mark and measure how deeply and extensively students will be expected to demonstrate and communicate their learning. DOK descriptors in the CRMs provide content-specific examples that illustrate how students might move towards deeper understanding with more complex or abstract content.

In 2005, Hess combined these two models for describing rigor and deeper learning. She saw that although related through their natural ties to the complexity of thought, Bloom's thinking levels and Webb's depth-of-knowledge levels differ in scope, application, and intent (Hess et al., 2009). The Hess Cognitive Rigor Matrix assists teachers in applying what cognitive rigor might look like in the classroom and guides test developers in designing test items and performance tasks. Content-specific descriptors in each of the Hess Cognitive Rigor Matrix's are used to categorize and plan for various levels of analysis and the mental processing required of assessment questions and learning tasks.

Hess saw that complexity resides not only in the demands of problem representation, but also in the levels of knowledge and application needed to formulate a problem solution. This framework helps in vividly connecting, yet clearly distinguishing, the two frameworks, allowing educators to examine the rigor associated with tasks that might seem at first glance comparable in complexity. Because cognitive rigor encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities, the Cognitive Rigor matrix can enhance instructional and assessment practices at the classroom level as well. (Hess et al., 2009).

Statement of the Problem

Assessments should serve to illustrate the proficiencies and deficiencies within students learning and to create an instructional plan for instruction and teaching methodology (Brown et al., 2014)**.** Most assessments operate from an understanding about how students learn, their cognitive developments and demand, as well as what skills students must learn. However, if assessments do not align to curriculum and cognitive demand, then this assessment tool will be deficient in helping to gauge student learning. Ultimately, annual yearly progress assessments such as PARCC are meant to show growth in learning. If it is not an accurate measure than achievements in important academic competencies will fall short (Brown et al., 2014)**.**

Prior research indicates that state achievement tests have not been measuring deeper learning to a large degree. Researches Yuan and Li argued that, "the Common Core State Standards (CCSS) initiative may increase the assessment of deeper learning nationwide" (Yuan & Vi-Nhuan, 2014). Over forty-five states have adopted the Common Core State Standards. These tests emphasize and claim deeper learning to a greater extent than other types of large-scale achievement tests. However, there has been no systematic empirical examination of the extent to which other widely used achievement tests emphasize deeper learning (Herman et al., 2014).

Current research on the factors influencing student outcomes and contributing to academic richness supports the concept that learning is optimized when students are involved in activities that require complex thinking and the application of (Hess et al., 2009). Even though test makers continue to focus on academic rigor and a standardized approach to assessing students learning, there is a lack of research that exists in gauging whether these tests truly measure higher order thinking and cognitive complexity within students. While PARCC claims that, “ PARCC cognitive complexity measures account for those things,” as well as, “other indicators of cognitive complexity (e.g., Bloom’s Taxonomy, Depth of Knowledge)” no tests have analyzed the test in regards to the extent of higher order thinking and depth of knowledge (PARCC, ETS, PEARSON, Assessment SIG Business Meeting , 2014).

Furthermore, while there have been multiple research studies that have analyzed PARCC tests for effectiveness, accessibility and complex tasks there have been none that have addressed the complexity of the exam in regards to higher order thinking and depth of knowledge (State of New Jersey Department of Education, 2016). As a result of this lack of research, more qualitative content analysis of the cognitive complexity of the PARCC exam is important. School administrators and personnel lack the empirical information necessary to make informed decisions about the PARCC exam. Educational Administrators might lack important information necessary to ensure quality assessments for all students.

Purpose of the Study

The purpose of this study was to determine how the language of written question prompts and activities in a publicly available online 10th Grade PARCC practice assessment associate with the language of higher order thinking found in literature, as represented on the Hess’ Rigor of Matrix. The research employed a mixed-method approach with qualitative content analysis and descriptive statistics. Specifically, the language from question prompts for tests on the 10th Grade PARCC Assessment test were analyzed and compared to the language associated with higher order thinking found in the research literature.

Research Questions

The study was grounded by an overarching research question: What are the types of thinking are assessed by the questions on 2019 PARCC practice tests in English Language Arts and Geometry in grades 10?

1. In what way(s) does the language of the questions on the English Language Arts section of 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 10 associate with the language that promotes higher-order thinking found in research literature?

2. In what way(s) does the language of the questions on the Geometry section of 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 10 associate with the language that promotes higher-order thinking found in research literature?

3. What is the distribution of thinking on the 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in English language arts and Geometry in Grade 10?

Methodology Overview

For the current study all 97 questions were analyzed from a grade 10 NJSLA Practice test. The test was comprised of 58 English Language Arts Questions and 39 Mathematics Questions. Two coders collected data by reviewing each selected question and aligning it with the language found in the various quadrants found within the Hess’ Cognitive Rigor Matrix. After all questions were aligned, the two coders compared the frequency and distribution of selected questions within the Hess’ Cognitive Rigor Matrix, focusing on the frequency of questions that fell into the category of higher level thus requiring higher order thinking skills.

Similarly, a focus was also on the frequency of questions that fall into the category of lower-level questions requiring students to simply recall or reproduce. This study is comprised of a mixed method research with an emphasis on quantitative statistics to explain the percentage of questions that fall within the cell of the Hess' Cognitive Rigor Matrix. The Hess' Cognitive Rigor Matrix and 10th Grade NJSLA Practice assessments were selected as the focal point of this analysis study.

Moreover, subsequent the grade level equivalency was selected predicated on the lack of specific research in high school 10th Grade PARCC Practice Tests. A further explanation of the coding scheme and procedure will be provided in Chapter III.

Conceptual Framework

Using Hess' Cognitive Rigor Matrix provides a theoretical framework to help guide the rationale for the research questions that I plan to investigate. Because no simple one-to-one correspondence relates Bloom's Taxonomy and depth of knowledge, Hess (2006b) superposed them. The resulting cognitive rigor (CR) matrix in Table 3 vividly connects, yet clearly distinguishes, the two schemata, allowing educators to examine the rigor associated with tasks that might seem at first glance comparable in complexity.

Although extending a pattern in mathematics may not look anything like distinguishing fact from opinion in English language arts, the two tasks reside in the same cell of the CR matrix and therefore share many common features: Both evoke similar thought processes and require similar instructional and assessment strategies.

The conceptual model will help to establish a structure that guides the query. Moreover, it builds upon previous research on current understandings of cognitive complexity. Webb's work highlights that both the content assessed in a test item and the intended cognitive demand, or the depth to which we expect students to demonstrate understanding of that content, is instrumental in aligning appropriate outcomes in assessments.

The cognitive rigor is a combined model developed using both Webb's Depth of Knowledge and Bloom's Taxonomy. It measures the depth and extent that instructional activity engages a learner's knowledge and thinking. The rationale for using Hess' matrix over Webb's Depth of Knowledge was that it offered more validity in determining cognitive complexity within assessment questions. Moreover, Hess' model has become the marker for evaluating current curriculum, as well as developing assessment items.

The 2019 NJSLA released practice tests reflect that of the items offered on the actual assessment of NJSLA. Because of the changes within NJSLA using the practice tests that reflect questions that will currently be asked forms the basis of the rationale for using these as for our sampling data.

Limitations

There might pose several limitations with this approach when interpreting the results from this study. One is treating cognitive demand as a fixed characteristic on a test item. Cognitive demand varies depending on the individual test taker's personal learning and background, the testing environment, and the skills being assessed on the test.

A further limitation might be the degree and extent that the sample assessments actually mirror the questions released from PARCC. There are usually differences and variations among released and sample items and the actual given assessment. Also, results are not generalizable to other tests or other versions of the PARCC test.

Finally, the evaluation of cognitive complexity within the assessments is a fixed framework. There are multiple ways in which one can evaluate cognitive complexity and higher order thinking. Alternative measures, might produce some variations in results.

Delimitations

There are several strategies that might be employed to ensure reliability and validity. The first is to use a framework to help guide evaluation of test questions. Research will support the validity in employing Hess' matrix within this research. It will serve as an instrument to externally validate our rational. The second is to code with multiple evaluators to ensure positive correlations on the established cognitive rigor framework. This will help to establish reliability to ensure that our answers and observations are consistent and aligned with the framework utilized.

Webb's depth of knowledge (Webb et al., 2005) was utilized as the conceptual framework for this study. Webb's DOK consists of four levels of knowledge including Level 1, recall, and Level 2, skills and concepts. These particular levels require basic knowledge recitation and comprehension. No complex thinking is present in DOK Levels 1 and 2. Webb's depth-of-knowledge Level 3, strategic thinking and complex reasoning, as well as Level 4, 10 extended levels of thinking, require students to reach deeper and think analytically and strategically. It is at DOK Levels 3 and 4 where researchers argue that complex thinking begins. This is in contrast to DOK Levels 1 and 2 that do not require this depth of thought. Being that the NJSLS have been adopted by the New Jersey State Board of Education, it is vital that these same standards are evaluated utilizing Webb's DOK levels to ensure that these standards include complexity, requiring high levels of complex thinking skills.

Definition of Terms

Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, synthesizing, and/or evaluating information gathered from, observation, experience, reflection reasoning, or communication as a guide to belief and action" (Su et al.,2016)**.**

Higher level cognitive questions Higher-order thinking questions are those that involve the learning of complex judgmental skills such as critical thinking and problem solving.

Higher order thinking are complex processes involving non-routinized thinking in which right and wrong answers cannot always be specified and in which complex reasoning and nuanced judgment may produce responses not previously encountered by an instructor" (Resnick, 1985, p. 9), specifically the skills metacognition, inference form context, decontextualization, information synthesis. Higher order thinking skills are thinking processes that require more than simple recall of facts (J. GORDON EISENMAN, 1995).

Lower-level cognitive questions are more basic and often asked mostly by teachers. Lower-order thinking skills are reflected by the lower three levels in Bloom's Taxonomy: *Remembering, Understanding, and Applying*. These types of questions ask students to recall and are generally fact based, closed questions that measure knowledge level.

Significance of the Study

Previous studies in school districts such as in New York City have applied Hess' framework to analyze levels of rigor then revise the rigor of questions using the Cognitive Rigor Matrix across subject areas (NYC department of education, 2018). Hess is mostly use to analyze curriculum not assessments. While other studies have measured certain complexities within test items there has not been a focus on this test, nor did they use established frameworks within their evaluation.

While there have been many studies that address the validity and complex thinking design embedded within the common core curriculum standards there are limited studies that evaluate the complex thinking demands within PARCC tests. The purpose for this qualitative content analysis study was to describe and compare the extent of cognitive complexity, as defined in Hess Cognitive Rigor Matrix, embedded in the Partnership for the Assessment of Readiness for College and Career (PARCC) practice tests for English Language Arts and Geometry.

With the ever-increasing push to, “close the achievement gap,” and to demand curriculum and standards that encourage complex thinking and rigor from students in order to develop today’s generation of students then assessments must align with these designs.

The Common Core State Standards were launched in 2009 by state leaders and governors from over 48 states with the promise of real-world learning goals and launched this effort to ensure that all students were prepared for college, career and life. As per the Common Core themselves, the standards were informed by teachers, content experts, states and leading thinkers, as well as feedback from the public (Initiative).

However, the implications behind standardization and national testing goes beyond instructional relevance. The focus on testing in the classroom overstresses the cognitive domain and does not take into account the other domains and methodologies that influences student outcomes. Students who fail to graduate from high school can experience lower life expectancy. In some populations it was four times lower (Hummer & Hernandez, 2015). Yet, national testing and educational decision-making is being driven that significantly impacts instructional decision making at the local level without accounting for factors that actually move outcomes.

Having a cognitive model that explicitly lays out the instructionally relevant, intermediate steps leading to expertise is the key to meaningful and useful alignment between the standards, curriculum, instruction, and assessment. This study should shed further light on the extent and degree to which complex thinking is embedded within PARCC.

Data Analysis Plan and Coding Scheme

The most appropriate way to organize the data are to organize it by PARCC sections and questions. To mirror the sections and questions will create a natural organization of analyzing each item within the test. With this type of organization, the concepts and data will be organized into a basic structure for my coding scheme.

Some themes and topic areas within the codes will be aligned with Webb's Depth of Knowledge and Bloom's Taxonomy. For instance, questions might be coded using the following verbiage: Remember/Recall, Understand/Skills, Understand/Thinking, Understand/Extended, Apply/Recall, Apply/Skills, Apply/Thinking, Apply/Extended, Analyze/recall, Analyze/Skills, Analyze/Thinking, Analyze/Extended, Evaluate/Thinking, Evaluate/Extended, Create/Recall, Create/Skills, Create/Thinking, Create/Extended.

This analytical approach will serve to address the extent that complex thinking is embedded in PARCC high school items. This approach uses an established theoretical framework to establish validity. Additionally, coding using multiple researchers will help to establish reliability. Ultimately, a hybrid in coding of pre-set and emergent codes will help to create flexibility in analyzing content items.

Chapter II: Review of Literature

Introduction

The purpose of this literature review was to evaluate the way(s) in which language found on the English Language Arts and Mathematics sections of the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 9 and 10 associated with existing literature regarding higher order thinking and the importance of questioning in developing higher order thinking skills. The research questions guide my review of the literature and encompass the following sections: The literature review will also present a review of definitions of higher order thinking in school curriculum and research literature. Additionally, this literature review identifies analyses of higher order thinking and its alignment with the Hess Cognitive Rigor Matrix.

Overview of Current Literature

The research in the area of PARCC's influence to inform classroom practices, and more importantly the extent that cognitive complexity is embedded in PARCC is very limited. The first part of the literature review centered on complex; higher order thinking resulted in a myriad of peer-reviewed literature. Much of the research expressed the need to develop complex analytical thinking within students and emphasized a need to develop 21st century skills.

The second part of the literature review focused on the common core standards and the correlation to cultivating developing higher order thinking. Finally, this review concluded with analyzing summative state assessments and their influence on curriculum, as well as demand on student thinking.

Literature Search Procedures

The peer-reviewed literature gathered for this review was found utilizing multiple online databases including ERIC, SAGE, and EBSCO, Academic Search Premier and ProQuest. Each component was individually searched for using key words such as PARCC, Hess Matrix, Webb's depth of knowledge, and complex thinking. In addition, Theories were researched including, Blooms Taxonomy, Webb's Depth of Knowledge, and Hess' Cognitive Rigor Matrix. The literature review included experimental, quasi-experimental, and meta-analysis.

Some specific works were sought due to their importance in other related studies. Non-peer-reviewed literature was also gathered from searching key terms and studies specifically related to complex analytical thinking and summative assessments.

Research used in this review included:

- Studies published within the last 20 years.
- Peer-reviewed research including dissertations and government reports,
- Non-peer reviewed surveys of skills desired by multinational corporations,
- Studies that focused on complex thinking,
- Dissertations,
- Peer- and non-peer-reviewed literature about the PARCC Claims,
- Frameworks utilized to assess learning standards/student learning, and
- Studies including the coding of specific learning standards,
- Works published by theorists regarding taxonomies, frameworks, and/or models for
- Higher order thinking.

Methodological Issues in Existing Studies of Complex Thinking in Assessments

There were various issues regarding the existing empirical research on complex thinking, and the complexity of the PARCC practice assessments. There includes a multitude of results surrounding the definition of higher order thinking, which yielded ambiguous results.

Furthermore, the terms around complex thinking, higher order thinking, cognitive complexity, rigor, critical thinking, strategic thinking, all were often interlaced in studies with no clarification in their distinctive influences in outcomes.

Although there have been many studies that address the validity and complex thinking design embedded within the common core curriculum standards there are limited studies that evaluate the complex thinking demands within PARCC tests. The Common Core Standards requires a high-level of cognitive demand; asking students to illustrate a level of understanding through the demonstration of content knowledge and academic skills. At the core of PARCC's goals, is to create a test that reflects and measures the Common Core State Standards, however limited research exists to illustrate a clear transference of the cognitive demands asked of the common core standards to a high level of cognitive demand on PARCC Assessments.

PARCC emphasizes that their "continuing design commitments reflect the Partnership's ambitions to meet these high expectations for next-generation, college and career readiness assessments. PARCC's claims and evidence statements directly reflect a commitment to measuring the 21st century cognitive competencies (PARTNERSHIP FOR ASSESSMENT OF READINESS FOR COLLEGE AND CAREERS)." However, there is no clear alignment on how PARCC is defining Cognitive Competencies and any correlating studies fail to make a continues definition of the characteristics for cognitive competencies. Yet, empirical evidence that demonstrates the efficacy of cognitive rigor outcome in PARCC remains elusive.

The PARCC consortia have adopted Evidence-Centered Design (ECD) as their approach to summative assessment development and validation (Mislevy & Haertek, 2007).

Evidence-Centered Design starts with the basic premise that assessment is a process of reasoning from evidence to evaluate specific claims about student capability. Student responses to assessment items and tasks provide the evidence for the reasoning process, and psychometric and other validity analyses establish the sufficiency of the evidence for substantiating each claim (Pellegrino et al., 2001).

In essence PARCC Exams can create artificial hierarchies that's true high order efforts are not accurately measured, as the frameworks for assessing test effectiveness are generated by non-peer reviewed outlets, which could indicate potential bias in methodologies and results. Once more, much of the literature around the PARCC test is centered on promoting and outlining the effectiveness of the exam with limited empirical references to validate effectiveness.

Review of Literature Topics

The purpose of this study with mixed-methods was to compare, analyze and examine the language of complex thinking embedded within the 2019 PARCC practice assessments in mathematics and language arts grades 9 and 10. That is, to what extent is the cognitive complexity, as defined in Hess Cognitive Rigor Matrix, embedded in the PARCC Practice Tests for English Language Arts and Mathematics in grades 9 and 10? The research will focus on the review of definitions of higher order thinking. Theoretical frameworks that assess cognitive complexity and cognitive rigor, like that of Hess' Cognitive Rigor Matrix, identified frameworks that are in alignment with the coding of PARCC assessments.

The conceptual framework implemented in this study was Hess' Cognitive Rigor Matrix. The Cognitive Rigor Matrix (CRM) vividly connects Webb's DOK and Bloom's Taxonomy, yet clearly distinguishes, the two schemata, allowing educators to examine the rigor associated with tasks that might seem at first glance comparable in complexity. Using the Cognitive Rigor Matrix can assist teachers in determining how thoroughly students understand the required content to be considered proficient (Hess et al., 2009a).

Purpose of Assessments in an Educational Setting

Assessment works within the educational system to help measure student learning and ability. However, it should not stand-alone. Pellegrino and Wilson argue that assessments are one of three-coordinated parts- curriculum, instruction, and assessment. Ultimately, "an assessment should measure what students are actually being taught, and what is taught should parallel the curriculum one wants students to master," (Pelligrino & Wilson, p. 5).

Moreover, assessments should serve to assist learning and the educational objectives of students. Assessments should serve to illustrate the proficiencies and deficiencies within students learning to create road maps for instruction and pedagogy. Ultimately, assessment is a tool designed to observe student's ability and skill and to produce date to this effect. Most assessments operate from a chain of reasoning about learning, cognitive demand, and what skills students must learn and at what level of progression in their development as a learner. With the ever-increasing push to, "close the achievement gap," and to demand curriculum and standards that demand complex thinking and rigor from students in order to develop today's generation of students then assessments must align with these designs.

My purpose for this qualitative content analysis study was to describe and compare the extent of cognitive complexity, as defined in Hess Cognitive Rigor Matrix, embedded in the Partnership for the Assessment of Readiness for College and Career (PARCC) practice tests for English Language Arts and Mathematics.

The Common Core State Standards were launched in 2009 by state leaders and governors from over 48 states with the promise of real-world learning goals and launched this effort to ensure that all students were prepared for college, career and life in an effort to help close the aforementioned achievement gap. As per the Common Core themselves, the standards were informed by teachers, content experts, states and leading thinkers, as well as feedback from the public (Initiative, 2018 Common Core State Standards).

The Common Core argues that, “Students who are able to master rigorous academic content, think critically and apply that knowledge to new problems, be conscientious and efficient in their learning, and communicate and collaborate with others are genuinely prepared for success” (PARCC). They argue that in addition to content knowledge and critical thinking skills, meta- cognitive and non –cognitive competencies are predictors in succeeding within college and in entry level jobs (PARCC).

As a result of NCLB, all 50 states and the District of Columbia have created state standardized tests for all children across the states of which the PARCC test resulted. “High stakes testing” was an effort to create an accountability tool. Adequate Yearly Progress (AYP) is a required statewide accountability system which requires each state to ensure that all schools and districts implement a “statewide accountability system mandated by the No Child Left Behind Act of 2001 which requires each state to ensure that all schools and districts make Adequate Yearly Progress". Ever since ESEA was passed, states and schools across the country have been working to improve its academics standards and assessments to ensure students graduate with the knowledge and skills most demanded by college and careers (Elementary and Secondary Education Act (ESEA), n.d.).

Complex Higher Order Thinking

The idea of developing instruction that progresses to engage complex higher order thinking has evolved over the years. Current research characterized complex thinking on a continuum of levels. For instance, lower levels of thinking often require identifying, memorizing or recalling information while higher levels ask learners to engage in more complex levels of thinking, such as synthesizing, evaluating and analyzing (Burns, 2017).

Cognitive complexity is defined as, “Cognitive complexity describes cognition along a simplicity-complexity axis. It is the subject of academic study in fields including personal construct psychology, organizational theory and human-computer interaction (Bell, 2004).” Meaning that it is cognition or the thinking skills that related to the organization of constructs in a given area or areas and their similarity.

Burleson and Caplan state that, “persons with highly developed systems of interpersonal constructs are better able than those with less developed systems to acquire, store, retrieve, organize, and generate information about other persons and social situations” (Burleson, & Caplan, 1998, p. 240). Additionally, cognitive complexity is strongly linked to individuals who are able to use cognitive complexity to better understand the world around them, as well as construct meanings quicker to the world around them (Dobosh, 2015).

Knauff and Wolf define complex cognition as, “…all mental processes that are used by individuals for deriving new information out of given information, with the intention to solve problems, make decision, and plan actions. The crucial characteristic of “complex cognition” is that it takes place under complex conditions in which a multitude of cognitive processes interact with one another or with other non-cognitive processes (Knauff & Wolf, 2010).” Meaning that in order to have complex cognition within learners, the task must draw on the learners’ previous constructs of material, interrelate and demand skills that are conducive to creating complex constructs.

Furthermore, in order to cultivate complex thinking curriculum needs to be adapted to include creative thinking and problem solving that progresses towards more rigorous cognitive tasks. For instance, developing curriculum that includes creating thinking and problem solving.

Higher Order Thinking

Some terms are often used interchangeable or interconnected with complex cognition. One term that often appears is Higher Order Thinking. It is widely used in the literature of cognition, however its definition tends to vary. Some see it as a type of thinking while others see it as consisting as activities that demand cognitive activities (Development Process, 2018). Thinking skills such as problem solving, decision making, conceptualizing, evaluating, synthesizing, and creative thinking require students to evaluate, plan, and monitor their thinking continuously (Eisenman Jr., 1995).

Higher order thinking is described as, "Complex processes involving non-routinized thinking in which right and wrong answers cannot always be specified and in which complex reasoning and nuanced judgment may produce responses not previously encountered by an instructor" (Eisenman JR., 1995 p. 22). Particularly skills in which learners must make inferences from context, decontextualize information, synthesize and use metacognition.

King et al emphasize higher order thinking as, “grounded in lower order skills such as discriminations, simple application and analysis, and cognitive strategies and are linked to prior knowledge of subject matter content. Appropriate teaching strategies and learning environments facilitate their growth as do student persistence, self-monitoring, and open-minded, flexible attitudes” (King, Goodson, & Rohani).

Thinking is a teachable and learnable skill and should not be reserved, as some misconstrue, for high achievers only. Teachers can expect, teach, and assess thinking skills for all students. Holding students accountable for higher-order thinking by using assignments and assessments that require higher order thinking in the forms of intellectual work and critical thinking increases student motivation as well as achievement (Brookhart, 2010, p. 12).

Stanley Pograw designer of the Higher Order Thinking Skills (HOTS) program specifically for educationally disadvantaged students, provided over ten years of research to the impact that higher order thinking assessment and instruction has on outcomes. The he designed worked on four kinds of thinking skills (Brookhart, 2010, p. 11):

(1) metacognition, or the ability to think about thinking;

(2) making inferences;

(3) transfer, or generalizing ideas across contexts; and

(4) synthesizing information.

In its 25-year history, the HOTS program has produced gains on nationally normed standardized tests, on state tests, on measures of metacognition, in writing, in problem solving, and in grade point average (Brookhart, 2010).

Many programs use some kind of taxonomy to help design and guide students thinking. One of the key characteristics of higher order thinking is that reasoning is required however, the mental infrastructure students need is less often discussed. Meaning, that how a student is primed for learning and the context they have in order to scaffold and achieve higher levels of thinking are not only harder to gauge but models for them are lacking (Brookhart, 2010).

What almost all of these frameworks for thinking and taxonomies have in common is that, as the number of elements (facts, concepts, statements, pieces of information) increases, and the number of relevant relationships among them increases, cognitive complexity increases (Brookhart, 2010). Students need to transfer their learning to contexts further and further from the one in which concepts were taught. Many curriculum documents and instructional materials use a cognitive taxonomy to ensure that higher-order thinking is taught and assessed, that students can transfer their knowledge to new situations.

While higher-order thinking is required when readers make complex inferences and integrate the text to their own constructs, not all challenging reading demands higher order thinking; only when students engage in higher-order thinking as they read complex texts and perform complex reading- related tasks are they using higher order thinking. Afflerbach and Cho argue that, "the most consequential assessments, high-stakes tests, are currently limited in providing information about students' higher- order thinking" (Afflerbach et al., 2015).

Critical Analytic Thinking

Another term that often appears when evaluating complex cognition is critical analytic thinking. Critical-analytic thinking requires, "that a student accept some level of uncertainty about a task and be open to multiple solutions; it implies a certain level of inquisitiveness and motivation to examine complex content deeply, well beyond simply recalling facts or restating answers" (Development Process, 2018). Critical analytic thinking combines content knowledge with analytic skills to assess the extent to which information supports or fails to support a particular proposition (Brown et al., 2014). However, cultivating critical analytic thinking and creating tasks that demand cognitive complexity is not necessarily the same thing. Deeper engagement within a test also does not necessarily facilitate cognitive rigor.

The Common Core State Standards, National policy, Next Generation Science Standards, as well as the National Assessment of Educational Progress frameworks all urge the need to assess a student's critical analytical thinking (Brown et al., 2014).

Developing a formative assessment that evaluates a student's critical analytical thinking skills poses many difficulties for large-scale assessments. The test content items, rubrics and overall construction must align not only with the standards and skills that are actually being taught in the classroom but also with the constructs of understanding in which a student operates. In developing large-scale assessments, test developers must judge the correlation between standards and assessment. The challenge is to create a test that matches content and cognitive demand. Despite these considerations, Herman et all posit that, "study after study has shown that current tests focus on lower levels of knowledge and application at the expense of those addressing deeper learning and high levels of cognitive demand" (Hermanet al., 2014).

Multidimensionality

It is important to note that many sources refer to multidimensional learning in both cultivating cognitive rigor and in creating meaningful curriculum. Moreover, PARCC claims that their test incorporates cognitive complexity measures that account for multidimensional learning (PARCC, ETS, PEARSON, Assessment SIG Business Meeting, 2014).

However, the term "multidimensional," is so broad in definition and scope that it can imply a myriad of many different concepts and considerations. Learning and cognitive development is itself multidimensional, which means that it occurs in many different dimensions, including biological, cognitive and socio-emotional (Takahashi, 2006).

Some models exist that attempt to facilitate this concept of multidimensional learning but often vary in how they approach this idea. For instance, Elferchichi et al believe that by providing personalized and well-adapted courses with the learner preferences, cognitive, emotional, social, and behaviorist level then multidimensional learning will occur (Elferchichi et al., 2007). Whereas, Abdelhamid discusses broaching multidimensional learning with a model that integrates different memory strategies to facilitate the learning process and is heavily dependent on illustrations and graphics (Abdelhamid, 1999). Learning can occur at any point on the spectrum of top down and bottom-up thinking orders. Overall, multidimensional learning should incorporate both teacher and learning that embeds complex thinking processes, student constructed meaning and a progression towards higher order thinking at various levels.

Although PARCC does incorporate multidimensional models of assessment, as opposed to one-dimensional assessments, it is not clear on how it caters to a multidimensional learner; perhaps, the claims that the assessment demands cognitive complexity from its test taker ties into some of these frameworks above. However, learning and problem solving on standardized tests do not adapt to learners' specific constructs and abilities and therefore cannot guarantee that cognitive complexity is occurring within the individual test taker. Such assessments do not take into consideration the zone of proximal development that a student is actually falling within. The zone of proximal development is "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving" (Eisenman, 1995 p.22).

Cognitive Rigor

Another term that often appears when evaluating complex cognition is cognitive rigor. Cognitive rigor encompasses the complexity of content, the engagement with the content as well as the scope of planned learning activities. Understanding how all of these components interact can enhance instructional and assessment practices (Hess, Jones, Carlock, & Walkup, 2009).

Cognitive rigor is considered a combined model developed by using existing models for outlining cognitive complexity. Specifically, it uses Bloom's Taxonomy and Webb's Depth-of-Knowledge levels. It is a, "concept that is marked and measured by the depth and extent students are challenged and engaged to demonstrate and communicate their knowledge and thinking" (Francis, 2017). Additionally, it helps to track and measure the complexity of student learning and instructional experiences (Francis, 2017). Ultimately, the Cognitive Rigor Matrix is intended to evaluate curriculum and lesson planning.

Assessment of Cognitive Domain Frameworks

Bloom's Taxonomy

The goal of using a cognitive taxonomy is to help students transfer their knowledge to new situations. The purpose of assessment of analysis, evaluation, or creation is to get information about the ways in which students use their knowledge and skills in novel situations (Brookhart, 2010).

Over the years, researchers and educator professionals have worked to develop frameworks to guide and assess student-learning demand within instructional tasks and activities. In 1956 educational psychologist, Dr. Benjamin Bloom developed a framework in order to promote higher forms of thinking in education. Rather than rote learning, he wanted a system that informed instruction and influenced higher order thinking (Clark, 2015).

The cognitive domain involves the development of intellectual skills and includes in six major categories in which cognitive processes are identified: knowledge, comprehension, application, analysis, synthesis, and evaluation (Clark, 2015).

Here are the authors' brief explanations of these main categories in from the appendix of *Taxonomy of Educational Objectives* (Armstrong, n.d.)

- Knowledge "involves the recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting."

- Comprehension “refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications.”

- Application refers to the “use of abstractions in particular and concrete situations.”

- Analysis represents the “breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between ideas expressed are made explicit.”

- Synthesis involves the “putting together of elements and parts so as to form a whole.”

- Evaluation engenders “judgments about the value of material and methods for given purposes.”

The multidisciplinary levels of cognitive development are illustrated in figure 1.

Figure 1

Business Simulations and Cognitive Learning, 2008

Learning Objective	Description of Learning	Assessment Process
Basic knowledge	Student recalls or recognizes information	Answering direct questions/tests
Comprehension	Student changes information into a different symbolic form	Ability to act on or process information by restating in his or her own terms
Application	Student discovers relationships, generalizations, and skills	Application of knowledge to simulated problems
Analysis	Student solves problems in light of conscious knowledge of relationships between components and the principle that organizes the system	Identification of critical assumptions, alternatives, and constraints in a problem situation
Synthesis	Student goes beyond what is known, providing new insights	Solution of a problem that requires original, creative thinking
Evaluation	Student develops the ability to create standards of judgment, weigh, and analyze	Logical consistency and attention to detail

Source: Bloom, Englehart, Furst, Hill, & Krathwohl (1959).

Revised Bloom's Taxonomy

In 2001, Bloom's Taxonomy was revised from a more static understanding of the taxonomy guiding objectives and learning outcomes to a more dynamic notion of "educational objectives." While this change may seem small it led further credence to the concept that learning happens on a spectrum and learners vacillate into different areas of the taxonomy throughout their learning experiences. It was this in mind that curriculum theorists and instructional researchers shifted the classification systems within Bloom's Taxonomy to verbs and gerunds.

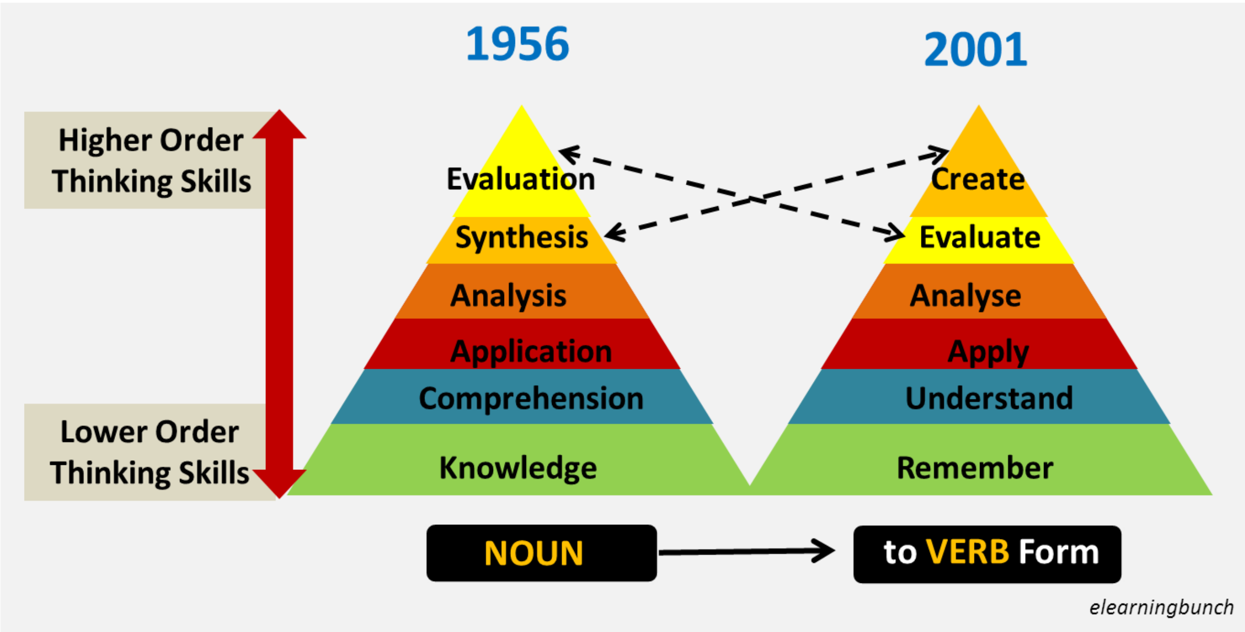
Lorin Anderson and David Krathwohl helped define the kind of thinking students are expected to demonstrate in order to respond to instructional tasks using Bloom's Taxonomy.

Ultimately, the split the cognitive domain of Bloom's Taxonomy into two dimensions that address: the knowledge dimension and the cognitive process dimension. The knowledge dimension addresses content and concepts, whereas the cognitive process dimension addresses the cognition of the learner. By doing so, they distinguished between the content that is being taught and what students must do while they are learning (Francis, 2017).

The two notable changes in these subcategories were the renaming of the knowledge category. Knowledge is an outcome of thinking, as this was on the lower end of the cognitive domain knowledge was renamed to remembering. Instructional researchers posited that thinking was too hard to further categorize and was too broad of a verb to replace with knowledge. The second notable change was comprehension and synthesis. These terms were replaced with understanding and creating respectively. They argued that this, "better reflected the nature of thinking defined in each category," (Marzano & Kendall, 2007).

Figure 2

(Wilson, n.d.) Revised Bloom's Taxonomy



Even though Dr. Bloom was at the forefront of creating frameworks to guide student cognition, cognitive complexity and higher order thinking does not happen in a linear singular progression from recall to analysis. Instead, higher order thinking is multifaceted and happens on varying levels, as learners engage and develop constructs within the learning process.

The new levels that identify cognitive learning are as follows (arranged from lower to higher order thinking (Wilson, n.d.) higher order thinking book:

Remember involves recognizing or recalling facts and concepts.

Understand involves basic comprehension, understood in light of newer theories of learning that emphasize students constructing their own meaning. Processes in this category include interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.

- Apply means to execute or implement a procedure to solve a problem. Application-level problems still usually have one best answer.
- Analyze means to break information into its parts, determining how the parts are related to each other and to the overall whole. Processes include differentiating, organizing, and attributing. Multiple correct responses are still likely in analysis-level tasks.
- Evaluate means judging the value of material and methods for given purposes, based on criteria. Processes include checking and critiquing.
- Create means putting disparate elements together to form a new whole, or reorganizing existing elements to form a new structure. Processes include generating, planning, and producing.

Figure 3

Comparison of descriptors (Hess, Jones, Carlock, & Walkup, 2009)

Bloom's Taxonomy (1956)	Revised Bloom Process Dimensions (2005)
Knowledge Define, duplicate, label, list, memorize, name, order, recognize, relate, recall, reproduce, state	*Remember* Retrieve knowledge from long-term memory, recognize, recall, locate, identify
Comprehension Classify, describe, discuss, explain, express, identify, indicate, locate, recognize, report, restate, review, select, translate	*Understand* Construct meaning, clarify, paraphrase, represent, translate, illustrate, provide examples, classify, categorize, summarize, generalize, infer a logical conclusion (such as from examples given), predict, match similar ideas, explain, compare/contrast, construct models (e.g., cause-effect)
Application Apply, choose, demonstrate, dramatize, employ, illustrate, interpret, practice, schedule, sketch, solve, use, write	*Apply* Carry out or use a procedure in a given situation; carry out (apply to a familiar task) or use (apply) to an unfamiliar task
Analysis Analyze, appraise, calculate, categorize, compare, criticize, discriminate, distinguish, examine, experiment, explain	*Analyze* Break into constituent parts, determine how parts relate, differentiate between relevant and irrelevant, distinguish, focus, select, organize, outline, find coherence, deconstruct (e.g., for bias or point of view)
Synthesis Rearrange, assemble, collect, compose, create, design, develop, formulate, manage, organize, plan, propose, set up, write	*Evaluate* Judge based on criteria, check, detect inconsistencies or fallacies, judge, critique
Evaluation Appraise, argue, assess, choose, compare, defend, estimate, explain, judge, predict, rate, core, select, support, value, evaluate	*Create* Combine elements to form a coherent whole, reorganize elements into new patterns/structures, generate, hypothesize, design, plan, construct, produce for a specific purpose

While educators still draw on Bloom's Taxonomy to guide their teaching, the verbs associated with a particular learning level do not necessarily correlate to the demand that is actually required of the learner. Moreover, it might not be the actual cognitive demand on a test question or activity (Hess et al., 2009).

Webb's DOK

There have been various attempts to define and guide what constitutes higher-order thinking in the public high school curriculum. The advantages of cognitive frameworks are that it allows researchers to deconstruct and categorize curriculum standards according to expected levels of cognition or thinking. Depth of knowledge forms another important perspective of cognitive complexity. Webb's (1997; 2007) Depth of Knowledge (DOK) another framework utilized in educational curriculums to help guide thinking.

According to Webb (1997), Depth of Knowledge encompasses multiple dimensions of thinking, including the "level of cognitive complexity of information students should be expected to know, how well they should be able to transfer the knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas" (Maverick Education, 2018).

Depth of Knowledge (DOK) is a way to define and categorize cognitive complexity of curriculum standards and tasks. The "DOK level of an item does not refer to how easy or difficult a test item is for students" (Wyse & Viger, 2011, p. 188). The focus of DOK is on the cognitive complexity of required tasks or curriculum standards.

In 1997, Norman Webb developed a process of criteria for analyzing alignment between standards and standardized assessments. Webb's depth of knowledge established and evaluated the depth of complexity of student learning experiences. Webb differs from Bloom in that he differentiates the context in which students will transfer and use deeper knowledge and thinking. The learning experience or "rigor" is measure by how in-depth students are expected to thin and interact with what they have learned. Additionally, unlike Bloom's Taxonomy, his model does not scaffold in complexity (Francis, 2017).

Webb’s Depth of Knowledge was conducive in influencing test alignment to include content assessed in a test item and the intended cognitive demand or the depth to which test takes expect students to demonstrate the understanding of the content. Cognitive frameworks have become essential functions in education reform and at the state level in the development of standards and assessment alignment (Hess, Carlock, Jones, & Walkup, What exactly do “fewer, clearer, and higher standards” really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments, 2009).

Webb (1997) developed a process and criteria for systematically analyzing the alignment between standards and standardized assessments. Since then, the process and criteria have demonstrated application to reviewing curricular alignment as well. This body of work offers the Depth of Knowledge (DOK) model employed to analyze the cognitive expectation demanded by standards, curricular activities and assessment tasks (Webb, 1997). The model is based upon the assumption that curricular elements may all be categorized based upon the cognitive demands required to produce an acceptable response. Each grouping of tasks reflects a different level of cognitive expectation, or depth of knowledge, required to complete the task. It should be noted that the term knowledge, as it is used here, is intended to broadly encompass all forms of knowledge (i.e., procedural, declarative, etc.) (Maverick Education, 2018).

Each grouping of tasks reflects a different level of cognitive expectation, and the DOK level should be assigned based upon the cognitive demands required by the ascribed task. The complexity of the task as well as the conventional levels of prior knowledge for students at the grade level, and the mental processes used to satisfy all relate to the various levels found in Webb’s framework Maverick Education, 2018).

The cognitive demands of many state accountability tests are analyzed with Webb's (2002) Depth of Knowledge levels. Webb uses four levels to classify the level of thinking required to do various cognitive activities (Hess et al., 2009; Webb, 1997):

DOK- 1 (Recall and Reproduction) - Curricular elements that fall into this category involve basic tasks that require students to recall or reproduce knowledge and/or skills. The subject matter content at this particular level usually involves working with facts and terms. In this level, there is little extended processing or thinking. At this level a learner, does not need to explore or synthesize to find an answer. They either can recall it or not.

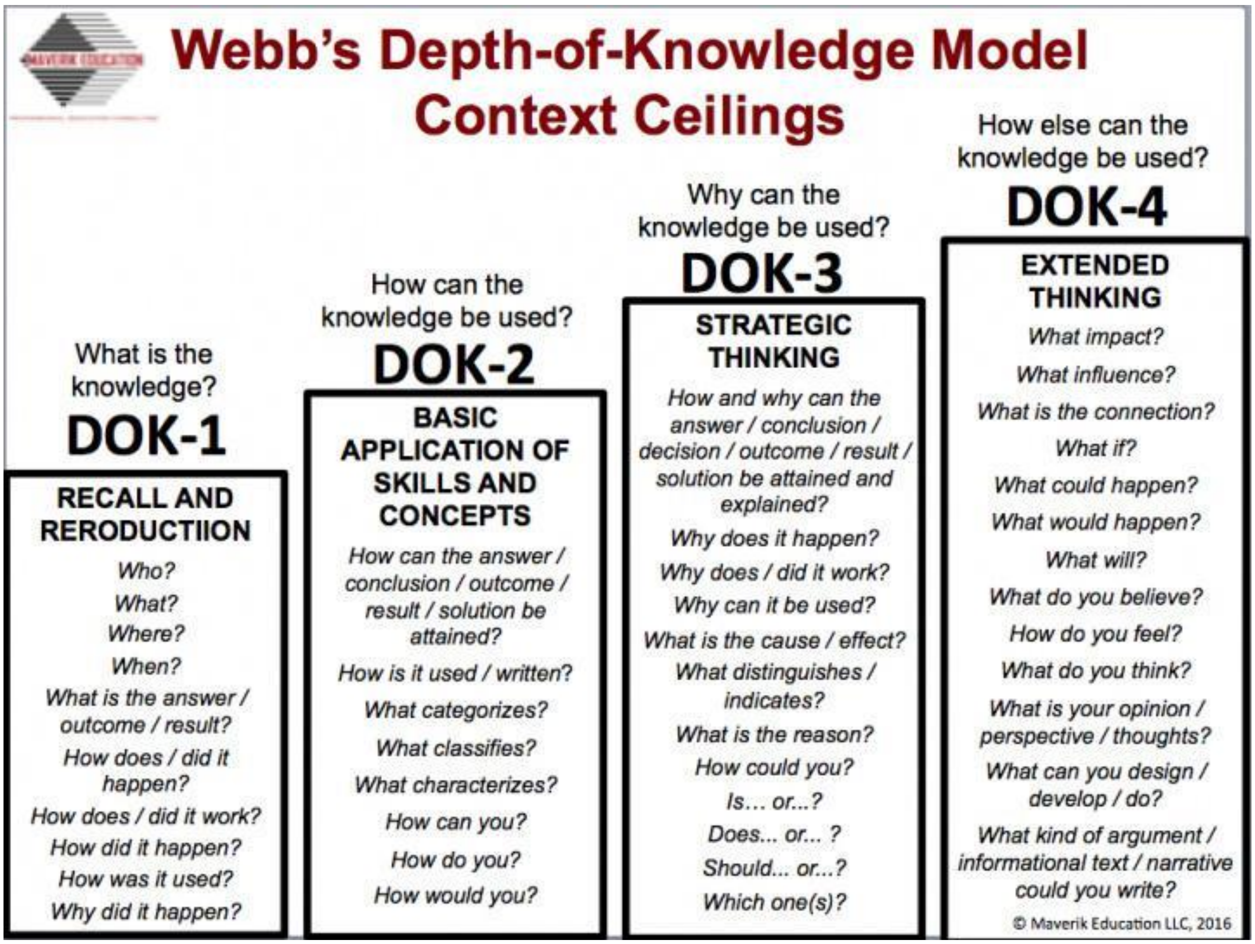
DOK- 2 (Skill and Concept) - This level includes the engagement of some mental processing beyond recalling or reproducing a response. This level generally requires students to contrast or compare people, places, events and concepts; convert information from one form to another; classify or sort items into meaningful categories. For instance, learners might describe or explain issues and problems, pattern, cause and effect, significance or impact, relationships, points of view or processes.

DOK- 3 (Short Term Strategic Thinking) - Items falling into this category demand a short-term use of higher order thinking processes, such as analysis and evaluation, to solve real-world problems with predictable outcomes. Learners explaining their reasoning is an essential characteristic within this level. The expectation established for tasks at this level tends to require a cross section of knowledge and skill from multiple subject-matter areas to carry out processes and reach a solution in a project-based setting.

DOK- 4 (Extended Thinking) - Curricular elements assigned to this level demand extended use of higher order thinking processes such as synthesis, reflection, assessment and adjustment of plans over time. Students are engaged in conducting investigations to solve real-world problems with unpredictable outcomes. Employing and sustaining strategic thinking processes over a longer period of time to solve the problem is a key feature of curricular objectives that are assigned to this level. Key strategic thinking processes that denote this particular level include: synthesize, reflect, conduct, and manage.

Figure 4

Webb's Depth of Knowledge Model Context Ceiling (Maverick Education, 2018).



Research theorists argue that effective schooling is based on four overarching criterion the educational environment, curriculum, instruction and assessment (Roach, Elliot, & Webb, 220). The degree to which these elements work together towards student learning is alignment – and the foundation of standards-based education reform. Alignment is the extent "to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do” (Webb, 2002, p. 1).

The development and implementation of large-scale assessment programs represent one approach to aligning classroom instruction with state curriculum standards. Webb’s model had been often referenced and utilized to evaluate the alignment between standards and assessment for language arts, mathematics, social studies and science in more than 10 states. These states have utilized this framework to modify assessment, alter standards and verify the extent to which these documents are directed towards the common goal of effective schooling and instruction (Roach et al.).

Hess’ Cognitive Rigor Matrix

Although related through their natural ties to the complexity of thought, Bloom's Taxonomy and Webb’s depth-of-knowledge differ in scope and application. Using these two frameworks, Carl Hess developed the Cognitive Rigor of Matrix. Bloom’s Taxonomy categorizes the cognitive demand and skilled needed for learners to perform a task whereas, Webb’s DOK relates to the depth of understanding which manifests from the skills required to complete the task. By creating a matrix of the two both the thinking processes and the depth of content knowledge can be better aligned and designed to direct implications in curricular design, lesson delivery, and assessment development and use (Karin Hess, 2006).

Because cognitive rigor encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities, the CR matrix has significant potential to enhance instructional and assessment practices at the classroom level. Superimposing the two cognitive complexity measures produces a means of analyzing the emphasis placed on each intersection of the matrix in terms of curricular materials, instructional focus, and classroom assessment.

The Cognitive Rigor matrix (CR) enables educators to examine the depth of understanding required for different tasks that might seem at first glance to be at comparable levels of complexity. Finally, the CR matrix allows educators to uniquely categorize and examine selected assignments/ learning activities that appear prominently in curriculum and instruction.

In two large-scale studies of the enacted (or taught) mathematics and English language arts curricula, teachers from 200 Nevada and Oklahoma public schools submitted over 200,000 samples of student work, which encompassed homework samples, tests, quizzes, and worksheets, completed during the period from February – May, 2008 (The Standards Company LLC, 2008a, 2008b as cited by Hess, et al, 2009a). Using the CR Matrix affords educators the opportunity to properly analyze curriculum and assessment for cognitive rigor.

Results for this study indicated that the majority of English Language Arts assignments correlated to the (DOK-2, Bloom-2) cell of the Cognitive Rigor. Mathematics assignments, sampled correlated to the (DOK-1, Bloom -3) cell to a greater extent. It was through this analysis that educators were able to determine that these types of assignments would not prepare students for non- routine applications or transfer of the same mathematics skills (Hess, Carlock, Jones, & Walkup, 2009).

The Cognitive Rigor Matrix enables educators the opportunity to properly analyze curriculum and assessment for cognitive rigor. They can then provide students with cognitively appropriate instruction. One conclusion the researchers have drawn from this work is that both measures of cognitive complexity can serve useful purposes in education reform at the state level (standards development and large-scale assessment alignment) and at the school and classroom levels (lesson design and teaching and assessment strategies). Ensuring that curriculum is aligned to “rigorous” state content standards is, in itself, insufficient for preparing students for the challenges of the twenty-first century.

Below is Hess’s Cognitive Rigor Matrix:

Figure 5.

Hess’ Rigor of Matrix *(Hess, Jones, Carlock, & Walkup, 2009)*

Table 2: Hess' Cognitive Rigor Matrix with Curricular Examples: Applying Webb's Depth-of-Knowledge Levels to Bloom's Cognitive Process Dimensions

Bloom's Revised Taxonomy of Cognitive Process Dimensions	Webb's Depth-of-Knowledge (DOK) Levels			
	Level 1 Recall & Reproduction	**Level 2 Skills & Concepts**	**Level 3 Strategic Thinking/ Reasoning**	**Level 4 Extended Thinking**
Remember Retrieve knowledge from long-term memory, recognize, recall, locate, identify	Recall, recognize, or locate basic facts, ideas, principles Recall or identify conversions: between representations, numbers, or units of measure Identify facts/details in texts			
Understand Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion (such as from examples given), predict, compare/contrast, match like ideas, explain, construct models	Compose & decompose numbers Evaluate an expression Locate points (grid/, number line) Represent math relationships in words pictures, or symbols Write simple sentences Select appropriate word for intended meaning Describe/explain how or why	Specify and explain relationships Give non-examples/examples Make and record observations Take notes; organize ideas/data Summarize results, concepts, ideas Make basic inferences or logical predictions from data or texts Identify main ideas or accurate generalizations	Explain, generalize, or connect ideas using supporting evidence Explain thinking when more than one response is possible Explain phenomena in terms of concepts Write full composition to meet specific purpose Identify themes	Explain how concepts or ideas specifically relate to other content domains or concepts Develop generalizations of the results obtained or strategies used and apply them to new problem situations
Apply Carry out or use a procedure in a given situation; carry out (apply to a familiar task), or use (apply) to an unfamiliar task	Follow simple/routine procedure (recipe-type directions) Solve a one-step problem Calculate, measure, apply a rule Apply an algorithm or formula (area, perimeter, etc.) Represent in words or diagrams a concept or relationship Apply rules or use resources to edit spelling, grammar, punctuation, conventions	Select a procedure according to task needed and perform it Solve routine problem applying multiple concepts or decision points Retrieve information from a table, graph, or figure and use it solve a problem requiring multiple steps Use models to represent concepts Write paragraph using appropriate organization, text structure, and signal words	Use concepts to solve non-routine problems Design investigation for a specific purpose or research question Conduct a designed investigation Apply concepts to solve non-routine problems Use reasoning, planning, and evidence Revise final draft for meaning or progression of ideas	Select or devise an approach among many alternatives to solve a novel problem Conduct a project that specifies a problem, identifies solution paths, solves the problem, and reports results Illustrate how multiple themes (historical, geographic, social) may be interrelated
Analyze Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, find coherence, deconstruct (e.g., for bias or point of view)	Retrieve information from a table or graph to answer a question Identify or locate specific information contained in maps, charts, tables, graphs, or diagrams	Categorize, classify materials Compare/ contrast figures or data Select appropriate display data Organize or interpret (simple) data Extend a pattern Identify use of literary devices Identify text structure of paragraph Distinguish: relevant-irrelevant information; fact/opinion	Compare information within or across data sets or texts Analyze and draw conclusions from more complex data Generalize a pattern Organize/interpret data: complex graph Analyze author's craft, viewpoint, or potential bias	Analyze multiple sources of evidence or multiple works by the same author, or across genres, or time periods Analyze complex/abstract themes Gather, analyze, and organize information Analyze discourse styles
Evaluate Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique			Cite evidence and develop a logical argument for concepts Describe, compare, and contrast solution methods Verify reasonableness of results Justify conclusions made	Gather, analyze, & evaluate relevancy & accuracy Draw & justify conclusions Apply understanding in a novel way, provide argument or justification for the application
Create Reorganize elements into new patterns/structures, generate, hypothesize, design, plan, construct, produce	Brainstorm ideas, concepts, or perspectives related to a topic or concept	Generate conjectures or hypotheses based on observations or prior knowledge	Synthesize information within one source or text Formulate an original problem, given a situation Develop a complex model for a given situation	Synthesize information across multiple sources or texts Design a model to inform and solve a real-world, complex, or abstract situation

5 2009 © Hess, Carlock, Jones, & Walkup. This article is currently pending publication. Permission to reproduce is given when authorship is fully cited.

The Partnership for Assessment for Readiness for College and Careers

The Common Core State Standards Initiative Website claims, “The Common Core focuses on developing the critical-thinking, problem-solving, and analytical skills students will need to be successful" (Burns, 2017). Brown et al, urges that, "having a cognitive model that explicitly lays out the instructionally relevant, intermediate steps leading to expertise is the key to meaningful and useful alignment between the standards, curriculum, instruction, and assessment (Brown et al., 2014).

Since the onset of the Common Core State Standards Initiative numerous studies have been employed to evaluate the alignment between standards and assessment and to incorporate judgments about the quality of the match between both content and cognitive rigor (Herman et al., 2014).

Researchers, Yuan and Li, for example, evaluated the complexity demands of released test items using Norman Webb's four-point depth of knowledge framework. Most questions were at a level of a 1 or 2. Overall, there were disparaging results from many states in regards to alignment between cognitive demand and expectations of their standards (Yuan & Vi-Nhuan, 2014).

The Smarter Balanced Assessment Consortium adopted the concept of Cognitive Rigor and the Hess Matrix in 2012 to measure the rigor of test items for the Next Generation of Assessments and used this matrix to evaluate the rigor outlined in the Common Core State Standards. Since its onset many states and districts utilize the Hess Matrix in order to align the state's large-scale assessments or to revise existing standards to achieve higher cognitive levels for instruction (Hess et al., 2009).

Overall, Hess' work suggests that, "There is ample evidence that the changes involved in teacher practices will raise the scores of students on normal conventional tests" (Karin Hess, 2006). The Common Core standards reflect the fact that inquiry and critical thinking are instrumental components in cultivating a student critical analytical thinking and preparing them for college and career (Hess, et al., 2012)**.**

In "Content Specifications for the Summative Assessment of the Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects," as well as "Content Specifications or the Summative assessment of the *Common Core State Standards for Mathematics,"* professional researched and educators, including Karin Hess, evaluated claims derived from Common Core State Standards as a resource to assist with the policy decision regarding the adoption of claims about student performance. Part of the development process began for this documented included an in-depth analysis of each standard in the CCSS document in every strand, at every grade level: Both the content and implied cognitive demand of each standard was analyzed. Overall, the consortium's findings were that most of the Common Core State standards assessed had a variety of selected response critical analytical thinking skills involved (Hess, et al., 2012).

The Common Core State Standards arguably have come closer to facilitating higher order thinking than other predecessors. However, the complexity of the standards does not just reside within the claim of the standard alone, but also in the ways in which that standard is employed within the instructional and assessment tasks. One of the main goals is the transfer of these skills within content areas. Studies have shown that, "students in thinking skills programs that teach the skills outside of a specific content area, such as reading, are able to transfer those skill back to the content areas" (Eisenman, 1995 p. 35). Therefore, it is not only important how the standards are being utilized and taught within the classroom but also how the assessment of those standards aligns with classroom instruction.

PARCC has also chosen an evidenced centered design (ECD) approach for the design and development of assessments. It is important to establish performance level descriptors (PLDs) at the initial stages of the assessment design and development work, because in ECD the PLDs will drive the validity and interpretative arguments (Kane, 1994).

The developers of the PARCC tests claim the tests measure students' readiness to master rigorous academic content at each grade level, think critically and apply knowledge to solve problems, and conduct research to develop and communicate a point of view (PARCC, 2019). The PARCC (2019) assessments claim to:

- Determine whether students are college and career ready or on track.
- Assess the full range of the Common Score State Standards, including standards that are difficult to measure.
- Measure the full range of student performance, including high and low performing students.
- Provide data during the academic year to inform instruction, intervention, and professional development.
- Provide data for accountability, including measures of growth.
- Incorporate innovative approaches throughout the system.

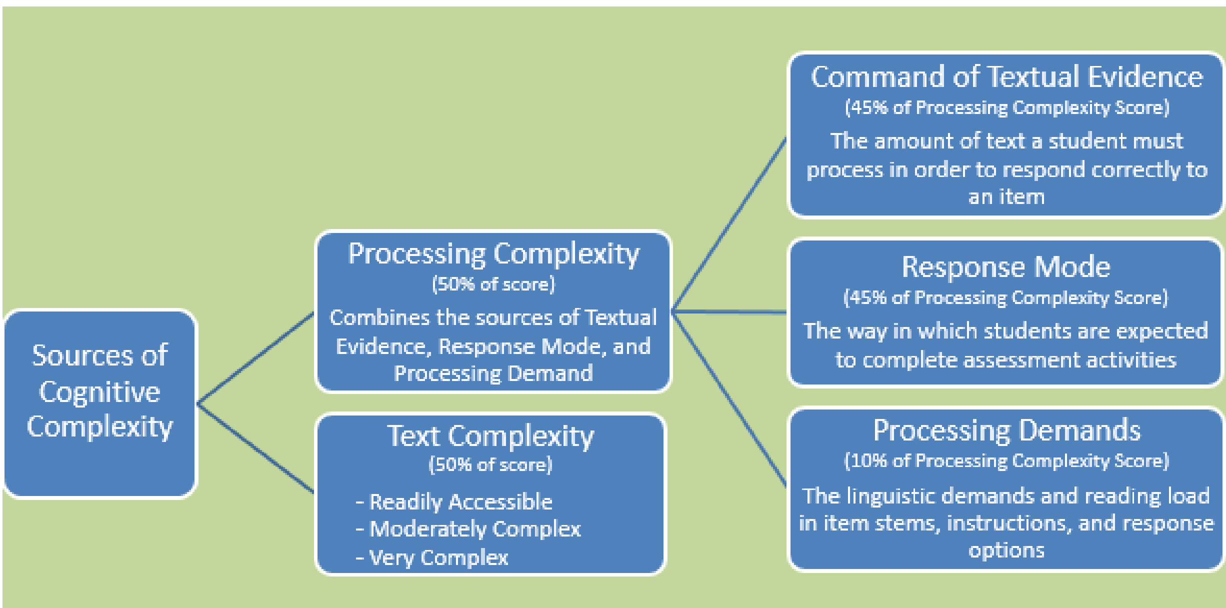
In 2016, PARCC switched to a single, end of year administration and in 2017, the PARCC Governing Board selected New Meridian Corporation as the management and content development vendor for the next phase of the PARCC assessment system (PARCC, 2019). The PARCC Assessments are based upon Evidence-Centered Design (ECD). Evidence-Centered Design is a systematic approach to test development. The design work begins with developing claims then transitions to evidence statements that were developed to describe the tangible things students could cite or reference, highlight or underline in a student work product that would help us prove our claims.

In English language arts the goals and uses of cognitive complexity are to:

- Provide a systematic, replicable method of determining item cognitive complexity
- Provide measurement precision at all levels of the test score scales

Figure 6.

Proposed sources of cognitive complexity in PARCC items and tasks: Language Arts (PARCC, 2019).



The PARCC Cognitive Complexity framework for ELA reflects the importance of text complexity as it relates to the CCSS, which indicates that 50 percent of an item's complexity is linked to the complexity of the text(s) used as the stimulus for that item. There are five sources of cognitive complexity as outlined by the framework. These include: processing complexity, text complexity, command of textual evident, response mode, and processing demands.

Consequently, to determine students' performance levels, it is critical to identify the pattern of responses when students respond to items linked to passages with distinct text complexities. To this end, PARCC claims to have developed a clear and consistent model to define text complexity and has determined to use three text complexity levels: readily accessible, moderately complex, or very complex (PARCC, 2019).

PARCC uses two components for determining text complexity for all passages:

a. Two quantitative text complexity measures (Reading Maturity Metric and Lexile) are used to analyze all reading passages to determine an initial recommendation for placement of a text into a grade band and subsequently a grade level.

b. Text Analysis Worksheets (New Meridian Corp, 2018), one for informational text and one for literary text, are then used to determine qualitative measures.

Items on the English Language Arts portion of the PARCC Test are designed to contribute to an understanding of how students "read closely to determine what the text says explicitly and to make logical inferences from it" and "cite specific textual evidence when writing or speaking to support conclusions drawn from the text" (Common Core State Standards, 2020).

.

Figure 7

Proposed sources of cognitive complexity in PARCC items and tasks: Mathematics (PARCC, 2019).

Performance Level	Level of Text Complexity[1]	Range of Accuracy[2]	Quality of Evidence[3]
5	Very Complex Moderately Complex Readily Accessible	Mostly accurate Mostly accurate Accurate	Explicit and inferential Explicit and inferential Explicit and inferential
4	Very Complex Moderately Complex Readily Accessible	Generally accurate Generally accurate Mostly accurate	Explicit and inferential Explicit and inferential Explicit and inferential
3	Very Complex Moderately Complex Readily Accessible	Minimally accurate Minimally accurate Generally Accurate	Explicit and inferential Explicit and inferential Explicit and inferential
2	Very Complex Moderately Complex Readily Accessible	Inaccurate Minimally accurate Minimally accurate	Explicit Explicit and inferential Explicit and inferential

For multimedia texts, qualitative judgments from one or both of the "optional" categories in the Complexity Analysis Worksheet will be combined with judgments in the other categories to make a holistic determination of the complexity of the material.

For the purposes of the PARCC Mathematics assessments, the major content in a grade/course is determined by that grade level's major clusters as identified in the PARCC Model Content Frameworks v.3.0 for Mathematics. Tasks on PARCC assessments provide evidence for this claim will sometimes require the student to apply the knowledge, skills, and understandings from across several major clusters. The PARCC Consortium claims that at each grade level, there is a range in the level of demand in the content standards--from low to moderate to high complexity. Within Mathematical Content, complexity is affected by (PARCC, 2019):

Low Complexity

- Items at this level primarily involve recalling or recognizing concepts or procedures specified in the Standards.

High Complexity

- High complexity items make heavy demands on students, because students are expected to use reasoning, planning, synthesis, analysis, judgment, and creative thought. They may be expected to justify mathematical statements or construct a formal mathematical argument.

The overall goals and uses of cognitive complexity in Matt are:

- Provide a systematic, replicable method of determining item cognitive complexity

  Provide measurement precision at all levels of the test score scales

- The PARCC assessments for mathematics will involve three primary types of tasks: Type I, II, and III.
- Each task type is described on the basis of several factors, principally the purpose of the task in generating evidence for certain sub claims.

Task Type	Description of Task Type
I. Tasks assessing ***concepts, skills and procedures***	• Balance of conceptual understanding, fluency, and application • Can involve any or all mathematical practice standards • Machine scorable including innovative, computer-based formats • Sub-claims A and B
II. Tasks assessing ***expressing mathematical reasoning***	• Each task calls for written arguments / justifications, critique of reasoning, or precision in mathematical statements (MP.3, 6). • Can involve other mathematical practice standards • May include a mix of machine scored and hand scored responses • Sub-claim C
III. Tasks assessing ***modeling / applications***	• Each task calls for modeling/application in a real-world context or scenario (MP.4) • Can involve other mathematical practice standards • May include a mix of machine scored and hand scored responses • Sub-claim D

Figure 8. Proposed sources of cognitive complexity in PARCC items and tasks: mathematics (PARCC, 2019).

The PARCC Cognitive Complexity framework for Math reflects that grade level's major clusters as identified in the PARCC Model Content Frameworks. There are six sources of cognitive complexity as outlined by the framework. These include: content complexity, processing complexity, practices complexity, stimulus material, response mode and processing demand.

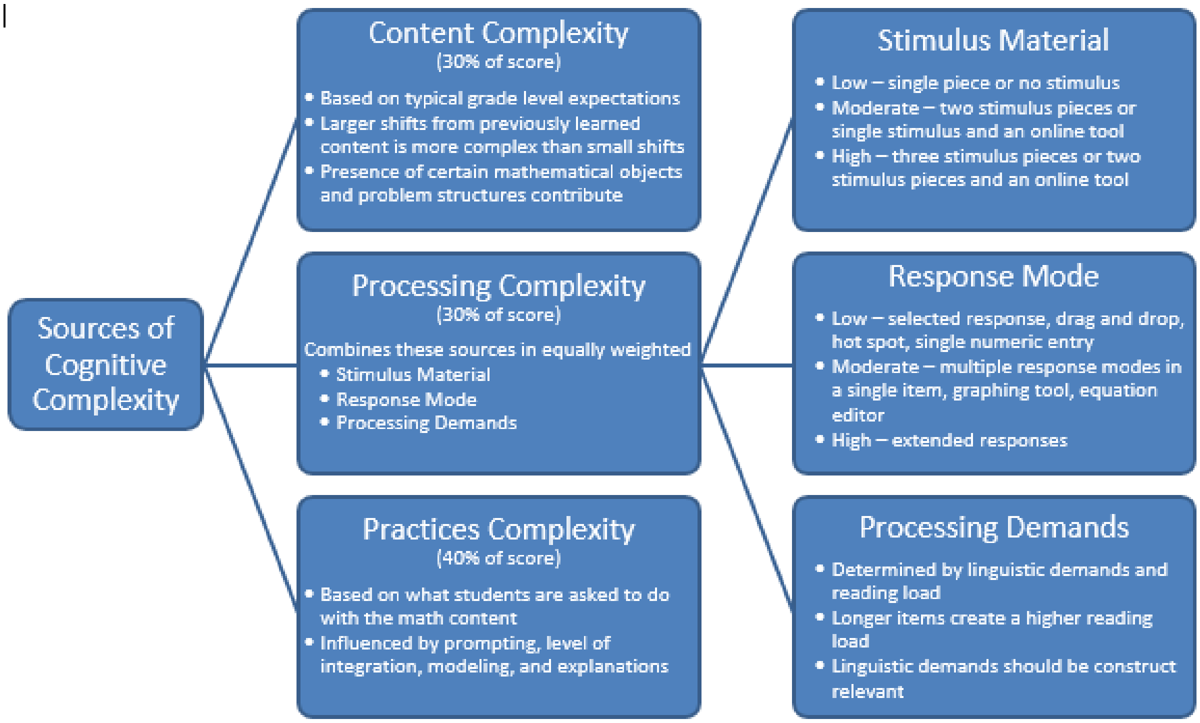


Figure 9. Proposed sources of cognitive complexity in PARCC items and tasks: mathematics (PARCC, 2019).

However, the PARCC Assessment has been dropped by many states Partnership for Assessment of Readiness for College and Careers. The test itself has received criticism not only for its length, for its technical glitches and for efforts by its test publisher but also on several cheating scandals surrounding the exam (Brown E., 2015).

Some communities have come to see PARCC as a symbol of federal overreach as PARCC receive millions of dollars to develop new tests aligned with the Common Core State

Standards, and yet is not a federal program (Brown E., 2015). Educators have criticized that the PARCC exam represents the overemphasis on standardized testing in America's schools.

Moreover, according to the technical manuals published by the creators of standardized assessments, none of the tests currently in use to judge teacher or school administrator effectiveness or student achievement have been validated for those uses. For example, none of the PARCC research, as provided by PARCC, addresses these issues directly. The tests are simply not designed to diagnose learning. They are simply monitoring devices, as evidenced by their technical reports.

Empirical evidence simply does not support the use of one-size-fits-all curriculum standards and high stakes testing as effective tools to improve the education and life outcomes of students. Nor does the evidence support the concept college ready skills or the skills that inspire creativity, citizenship and an overall "well rounded" student who will achieve can be achieved through standardization.

One claim made often to support PARCC testing is that New Jersey has had curriculum standards and state testing for a long time, and that the idea behind PARCC testing is not new. However, those original standards and tests were later deemed ineffective and of low quality by the NJDOE, the same organization that mandated the original standards and tests (Tienken, 2015).

Another claim that lacks empirical support is that the results from the PARCC tests will be diagnostic and tell us important things about student learning and the quality of the teaching that our children receive. To diagnose a student's achievement at the individual level, of any one skill, the test results must have reliability figures around .80 to .90. To attain that level of reliability there must be about 20-25 questions per skill (Frisbie, 1988; Tanner, 2001). Yet, the

PARCC tests do not have enough questions to diagnose student achievement at the individual level in any of the skills or standards. Thus, any "diagnostic" decisions made from PARCC results about a student's understanding of specific standards will be potentially flawed (Tienken, 2015).

Furthermore, PARCC tests are aligned to the Common Core State Standards. Those standards mandate knowledge and skills that are not much different than students have received for the last 150 years (Tienken, 2015). These standards have not been realigned or revisited to include many 21st century thinking and learning skills. We are moving away towards non-routine jobs and towards jobs that are ever demanding of critically thinking and problem solving. Whether states trying to measure proficiency or growth, standardized tests are not the answer.

Summative State Assessments

In 1965, in order to help decrease the gap in achievement throughout the United States, the federal government enacted the Elementary and Secondary Education Act. This act sought to create equal footing through the access to adequate education. Part of this act encouraged high educational standards, as a means of holding school districts accountable for the curriculum and instruction. They needed a system though in order to uphold and enforce these standards. As a result, the accountability system known as the Adequate Yearly Progress (AYP) was enacted. It is a statewide accountability system that ensures that all schools have made progress and have been working to improve its academic standards (OSPI, 2017).

Since then a number of federal and statewide initiatives have been enacted to carry out this federal mandate. In 2010, the PARCC consortium was awarded the Race to the Top assessment funds by the U.S. Department of Education to help develop an assessment to measure Common Core Standards. PARCC has led the movement towards creating Common Core State

Standards in English language arts/literacy and mathematics. They consulted with over 200 educators and administrators to help develop this assessment (OSPI, 2017).

PARCC attempted to design a test that considered multiple sources of complexity. It pushed beyond traditional assessment in their item design and evaluation. PARCC's sources of cognitive complexity have three levels of complexity. (Hess et al., 2009). PARCC argues that, "testing complexities mimic the rigor of the standards within the common core" (Brown, et al., 2014)**.**

However, statewide participation has been declining since it first was enacted. Over time, more and more states have withdrawn from the PARCC test. As of May 2016, only six stated plan to give this assessment; two of which plan on withdrawing in the 2018 (Jochim & McGuinn, 2016).

General Findings

The research in the area of developing complexity measures in PARCC, as well as the correlation to higher order thinking produced various peer-reviewed articles.

In the study, "Measuring Deeper Learning Through Cognitively Demanding Test Items Results from the Analysis of Six National and International Exams." researchers Yuan and Vi-Nhuan analyzed six national and international exams to assess how cognitively demanding these exams were. While they treated cognitive demand as a fixed characteristic of a test item, however they acknowledge that part of cognitive demand is the students' interaction and experience with the question based on their own constructs. Moreover, they based their study on already released items and cognitive demand might vary year to year based on the actually test design (Yuan & Vi-Nhuan, 2014, p. 42).

They found that although prior research indicates that state achievement tests have not been measuring deeper learning to a large degree that the Common Core may actually increase the assessment of deeper learning. In their study they examined how these statewide assessments measured deeper learning. Overall, the six benchmark tests demonstrated greater cognitive demand than did the state achievement tests in both subjects. Their study concluded that, “the average share of items rated at or above DOK level 3 was about 15 percent for mathematics and 40 percent for ELA across the six benchmark tests, compared with 2 percent for mathematics and 20 percent for ELA across the 17 state achievement tests included in an earlier study,” (Yuan & Vi-Nhuan, 2014 p.3).

They found that the majority of test items were coded at lower levels of cognitive demand (Yuan & Vi-Nhuan, 2014, p. 40).They argue that, “the level of cognitive demand is not the ultimate driving factor for test design” (Yuan & Vi-Nhuan, 2014, p. 41). Assessments like PARCC tend to focus on one aspect of deeper learning and do not necessarily assess intrapersonal and interpersonal competencies that are needed for deeper learning (Yuan & Vi-Nhuan, 2014). They conclude with a need for benchmarks and state achievement tests to develop analytic frameworks to help integrate critical thinking and problem solving in each subject area.

In the study, “Evaluating the Content and Quality of Next Generation Assessments” they attempted to answer if assessment place a strong emphasis on the content outlined for college and career readiness and by the Common Core State Standards. They also looked at whether or not the range of thinking skills included a range of high order thinking skills (Doorey & Polikoff, 2016). They concluded that PARCC and Smarter Balanced assessments also contain a distribution of cognitive demand that better reflects that of the standards, when compared to ACT Aspire and MCAS.

In order to overcome limitations in other studies they created a new methodology designed to address tests that focus on college and career readiness. The methods were develop by experts at the National Center for the Improvement of Educational Assessment (NCIEA). They found that PARCC's ELA/Literacy assessment require a range of cognitive demand, and include a variety of item types. Overall, the study measured, "how well new tests measure the requisite content, knowledge, and critical skills at key grade levels and, in doing so, whether they sufficiently tap higher-order thinking skills," (Doorey & Polikoff, 2016).

In, "A Comparison of Complex Thinking Required by the Middle School New Jersey Student Learning Standards and Past New Jersey Curriculum Standards," research Burns looked at the demand on complex thinking in middle school student learning standards. He was compelled in his research because no empirical evidence currently existed regarding the DOK levels of the NJSLS in Grades 6–8 compared to the DOK levels contained in the NJCCCS at these same grade levels. He found that, "seventy-two percent (18% of DOK Level 1 and 54% of DOK level 2) of the Grades 6–8 mathematics NJSLS were rated at a DOK Level 1 and 2. Relatedly, 28% (24% of DOK 3 and 4% of DOK 4) of the Grades 6–8 mathematics NJSLS were rated at a depth-of-knowledge Level 3 and 4. This evidence suggests that the Grades 6–8 NJSLS in mathematics contain a vast majority (72%) of its standards falling in the lower recall (DOK Level 1) and skill/concept (DOK Level 2) categories," (Burns, 2017, p. 100). Overall, his findings indicate that if a curriculum is based on content standard that are low in cognitive complexity than it will be difficult for students to develop 21st century and college readiness (Burns, 2017).

Finally, in "Human Resources Research Council: Evaluation of the Content and Quality of the 2014 High School MCAS and PARCC Relative to the CCSSO Criteria for

High Quality Assessments” reviewed high school assessment, including PARCC. The HumRRO study found the PARCC ELA/literacy test to an appropriate mix of text types, and rigorous items. While in the mathematics portion of the test the cognitive demand of the test aligned with the standards. They asserted that the ELA test had a heavier emphasis on high order thinking skills than the standards. Reviewers concluded that higher level items should be more valued on the design of the test (Schultz et al., 2014).

Conclusion

The overall research suggests that despite challenges in developing statewide formative assessments the PARCC tests are an improvement on former assessments and represent a move in the right direction for educational tests in states. However, further frameworks and considerations need to be developed in order to make sure that these assessments are able to assess higher levels of complex analytical thinking. The design of an actual assessment is a challenging endeavor that must be guided by theory and research about cognition in context. (Development Process, 2018). As Eisenman asserts, "Complex processes involving non-routinized thinking in which right and wrong answers cannot always be specified and in which complex reasoning and nuanced judgment may produce responses not previously encountered by an instructor" (Eisenman, 1995 p. 65).

It is important to remember the functions that assessments play in an educational setting in regards to student’s competencies. Ideally, an assessment should measure what students are actually being taught. Hess et all emphasizes that, “research and theory would suggest that complexity of problem solving depends on the extent to which a problem is well or ill-defined, routine or demanding far transfer, and the amount of information and interactions that require processing as a student plans and monitors his/her solution processes” (Hess et al.,2009).

Theoretical Framework

Throughout the United States educational programs have used and referenced numerous theoretical frameworks to help guide higher order thinking and learning within students. One of the more popular frameworks that is still employed today is Bloom's Taxonomy. Bloom's focuses on six categories of thinking. The revised version of this framework focuses on verbs, as opposed to nouns that students should be enacting in that level.

Perhaps one of the more widely misconceived notions of this framework is that students do not scaffold from one lower level of thinking to the next higher. Instead, students throughout the course of a lesson or assignment will vacillate from lower to higher, or higher to lower. The idea that higher order thinking happens on a spectrum and not a tiered process is a concept that is often failed to be emphasized when referencing theoretical frameworks.

One of the more popular frameworks in helping to analyze the alignment of thinking processes involved in designated curriculum and assignments is Webb's (1997, 2002) DOK. Webb (1997) developed a process and criteria for systematically analyzing the alignment between standards and test items in standardized assessments. Webb's model categorizes assessment tasks by different levels of cognitive expectation, or depth of knowledge, required to successfully complete the task. Webb's helps to assist current methodologies for judging the alignment between standards and assessment routinely incorporate judgments about the quality of the match in both content and cognitive rigor. Depth of knowledge forms another important perspective of cognitive complexity. Probably the best- known work in the area of depth of knowledge is that of Norman Webb (1997, 1999). Webb's work has forced states to rethink the meaning of test alignment to include both the content assessed in a test item and the intended cognitive demand, or the depth to which we expect students to demonstrate understanding of that content. In other words, the complexity of both the content (e.g., simple vs. complex data displays; interpreting literal vs. figurative language) and the task required (e.g., solving routine vs. non-routine problems) are used to determine DOK levels (Hess, Carlock, Jones, & Walkup, 2009).

Hess (2004-2012) further articulated the model with content specific descriptions for use by classroom teachers and organizations conducting alignment studies. Because no simple one-to-one correspondence relates Bloom's Taxonomy and depth of knowledge, Hess (2006b) superposed them. The resulting cognitive rigor (CR) matrix in Figure 4 vividly connects, yet clearly distinguishes, the two schemata, allowing educators to examine the rigor associated with tasks that might seem at first glance comparable in complexity.

This study utilizes Hess' Cognitive Rigor Matrix as the framework to categorize the complexity of language used on PARCC practice test questions in 10th grade PARCC Practice Tests. Because cognitive rigor encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities, the CR matrix can enhance instructional and assessment practices at the classroom level as well Both Bloom's Taxonomy and Webb's depth of knowledge therefore serve important functions in education reform at the state level in terms of standards development and assessment alignment. Because cognitive rigor encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities, the CR matrix can enhance instructional and assessment practices at the classroom level as well.

Implementing Hess' Cognitive Rigor Matrix to guide analysis, this study requires coding and the comparison of various DOK and Bloom's Taxonomy levels in order to illicit important conclusions. This study ultimately aims to systematically examine the extent in which higher learning is embedded in the PARCC practice tests for 10th grade.

Using Hess' theoretical framework this study aims to compare the practices tests from Grade 10 English language arts and mathematics with the PARCC claims, while describing the level and distribution of higher order thinking. Since higher order thinking is not clearly defined and definitions vary in its classroom application the results of this study will help contextualize future analysis to the extent of cognitive complexity in the assessments compared to the state standards, and actual classroom application.

Chapter III: Methods

Introduction

The purpose of this mixed-method study focused on describing and categorizing the distribution of cognitive complexity, as defined by the language on the Hess' Cognitive Rigor Matrix, on 10th Grade 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) High School Practice Tests in Geometry and English Language Arts. I explored the topic due to lack of existing literature on the accuracy of the claims about increased higher level thinking on the high school PARCC tests. The following chapter describes the methodology used for this study.

The New Jersey Core Curriculum Content Standards (NJCCCS) were first adopted by the State Board of Education in 1997. The original NJCCCS and the corresponding state tests were not the result of extensive curricular research. Rather, the NJCCCS were imposed by the Whitman Administration as part of a lawsuit, known as Abbott versus Burke, over New Jersey's school funding formula. New Jersey's first set of mandated curriculum standards and high-stakes tests were created to satisfy legal and political mandates, not for educational reasons (Tienken, 2015).

With the enactment of the *No Child Left Behind Act of 200*1 (*NCLB*), New Jersey's statewide assessment system underwent further change. This federal legislation required that each state administer annual standards-based assessments to students in grades 3 through 8, and at least once in high school. The federal expectation was that each state would provide tests that were grounded in rigorous state content standards and that would assess student achievement in language arts literacy, mathematics and, at three benchmark grade levels, science.

In preparation for the new accountability system, many states joined the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium in the spring of 2010. (Department of Education State of New Jersey, 2016).

The Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in tenth grade were selected as the focal point of this analysis study due to the lack of existing literature on the level of complex thinking embedded in the assessment. Even though PARCC is being dropped by many states throughout the country, PARCC like tests and questions will still live on through shared licensing agreements between states and entities. For instance, Eno announced that instead of the Partnership for Assessment of Readiness for College and Careers, known commonly as PARCC, the tests will be called the New Jersey Student Learning Assessments (NJSLA) (Danzis, 2018). The following chapter describes the methodology, in detail, used for this study.

Research Questions

The study was grounded by an overarching research question: What are the types of thinking are assessed by the questions on 2019 PARCC practice tests in English Language Arts and Geometry in grades 10?

1. In what way(s) does the language of the questions on the English Language Arts section of 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 10 associate with the language that promotes higher-order thinking found in research literature?

2. In what way(s) does the language of the questions on the Geometry section of 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests

in Grades 10 associate with the language that promotes higher-order thinking found in research literature?

3. What is the distribution of thinking on the 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in English language arts and Geometry in Grade 10?

Research Design

This study used qualitative content analysis methods to describe and categorize distribution of replicable and valid inferences using high order thinking in (PARCC) high school practice tests for English Language Arts and Mathematics, based on Hess' Cognitive Rigor Matrix. *Mixed methods* approaches are used to gain a deeper understanding of measurable concepts, develop instruments, and validate theoretical models. The value of qualitative methods in mixed methods approaches is that they can reveal information that would have been impossible to uncover through quantitative methodologies alone.

Content analysis is a research tool used to determine the presence of certain words, themes, or concepts within some given qualitative data. Using content analysis, researchers can quantify and analyze the presence, meanings and relationships of such certain words, themes, or concepts.

The specific type of content analysis approach chosen for this research is the directed approach. This study uses a content analysis approach guided by a structured process using an existing theory or prior research. This case study design was utilized in this study as it provided the structural methods needs to study the cognitive complexity within PARCC assessments. Thus, this study aimed to identify the relationship between the questions provided in PARCC high school practice tests for English Language Arts and Geometry and its purpose to enhance cognitive rigor.

Methods

A qualitative content analysis method was employed for the first part of the study to code each of the PARCC assessment questions in language arts and mathematics Grades 10 based on pre-existing codes.

The purpose of content analysis is to organize and elicit meaning from the data collected and to draw realistic conclusions from it. Content analysis is a research method that provides a systematic and objective means to make valid inferences from verbal, visual, or written data in order to describe and quantify specific phenomena. Down-Wambolt underlines that “content analysis is more than a counting process, as the goal is to link the results to their context or to the environment in which they were produced” (Bengtsson, 2016).

The results of data analysis is to organize and elicit meaning from the data collected and draw realistic conclusions (Polit & Beck, 2006). Part of the analyzing process, included coding. Coding facilitates the identification of concepts around which the data can be assembled into blocks and patterns (Catanzaro, 1988). Codes were generated deductively for the design of this study. A deductive reasoning design, created a coding list before starting the analyzing process. Each identified meaning unit is labeled with a code, which should be understood in relation to the context. This procedure is recognized as the “open coding process” in the literature.

The coding protocol for each assessment question in each subject and grade level followed the procedures described by Mayring (2000). The coding team analyzed and coded the Grade 10 PARCC practice assessments in English language arts and Geometry based on the language found on the Hess Cognitive Rigor Matrix methodology (See Figure 4). The categories from the Hess Cognitive Rigor Matrix formed the foundation for the codes.

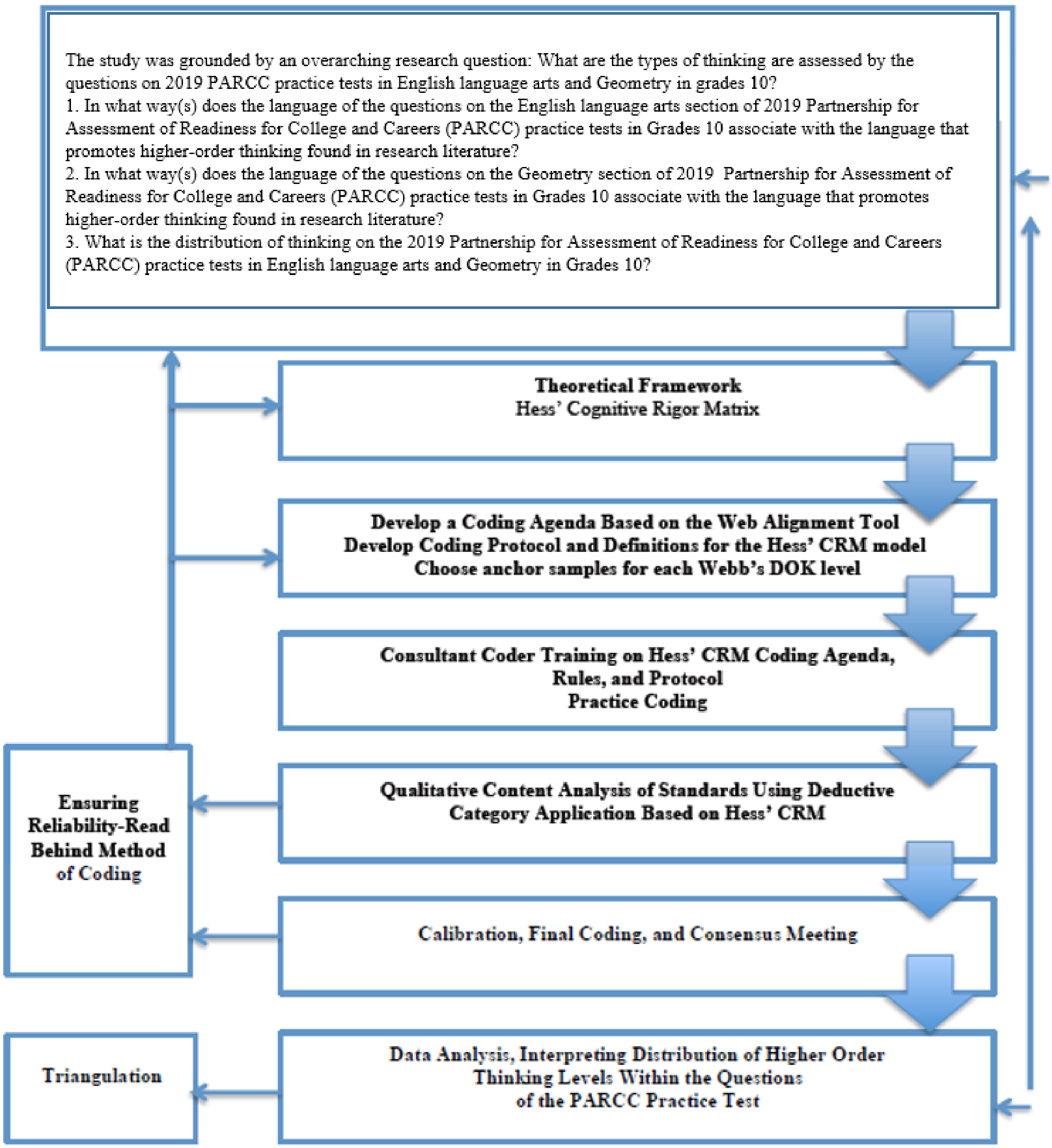
Deductive category application was utilized to connect the language from Hess' Cognitive Rigor Matrix to the language of the **97** English language arts and mathematics questions obtained from the PARCC practice tests. The results in deductive content analysis should be reported systematically and carefully, with particular attention paid to how connections between the data and results are reported (Elo, et al., 2014). An essential consideration when discussing the trustworthiness of findings from a qualitative content analysis is that there is always some degree of interpretation when approaching a text (Eloet al., 2014). As such, Hess' Rigor of Matrix was applied in this study to correlate comparable complexity found on the PARCC 10th grade ELA and Geometry Practice Tests. Complexity resides not only in the demands of problem representation, but also in the levels of knowledge and application needed to formulate a problem solution

The Cognitive Rigor Matrix (CRM) couples the higher order thinking of Webb's Depth of knowledge and the analysis of cognitive skills within tasks and assessments. The CRM directly correlates and clearly distinguishes, the two schemata, allowing educators to examine the rigor associated with tasks that might seem at first glance comparable in complexity (Hess, et al., 2012).

Quantitative methods were employed in the second part of this study. Specifically, Frequencies and descriptive statistics were utilized to describe the differences and similarities of complex thinking that exist in the language of the 10th grade ELA and Geometry PARCC practice assessment. I calculated the percentage of the questions that were categorized in each level of Hess' CRM based on the qualitative analysis of the language of the assessment questions.

Figure 10

Step model for deductive category application, adapted from Mayring (2000).



Description of Documents

The above listed components of quantitative content analysis will be for a qualitative oriented procedure of text interpretation. Deductive category application was employed for this study. Systematic text analysis is also employed within content analysis to develop qualitative procedures are methodological controlled. Those procedures allow a connection to quantitative steps of analysis if it seems meaningful for the analyst.

The English language arts and Geometry in the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 10 and Geometry documents analyzed in this study. Document were downloaded from the PARCC website on September 19, 2020. The practice tests in their entirety are a 114-page document that focuses on both Language Arts and Geometry in grades 10. The evidence statements describe the knowledge and skills that the assessment item/task elicits from students are derived directly from the Common Core State Standards for mathematics and language arts (the standards) (PARCC, 2019).

Data Collection

The data gathered was retrieved from a public website containing PARCC practice assessment information and various tools. The practice tests were freely available on the website. For this study, I only focused on Grade 10 English Language Arts and Geometry.

Coders

As part of this study a coding committee was established. Two coders were used. The first coder has been an educator for over 12 years in Grades 10 and 12, as well as Higher Level Education. The qualified second coder was asked to code and determine the proper placement of each assessment question utilizing Hess' Cognitive Rigor Matrix.

The second coder, was an educator and administrator for over 20 years and earned his doctorate in Educational Leadership in 2020. He has previous coding experience using Hess CRM since 2016. The coders followed and implemented the rules adapted from the Webb's Alignment Training Manual.

Coding Scheme

Hess' Cognitive Rigor Matrix contained sample performance tasks and example activities learners are asked to do in each of the cells of the matrix, intersecting Bloom's Taxonomy and Webb's Depth of Knowledge. The examples of performance tasks and activities provided a comprehensive list, which reduced the possibility of a question being coded incorrectly and increased the reliability among coders. Because Hess' Cognitive Rigor Matrix is designed as a grid, with Webb's Depth of Knowledge as the columns and Bloom's Taxonomy as the rows, a specific code was assigned to each cell to provide a more accurate and comprehensive coding scheme. The first number in the matrix described Webb's Depth of Knowledge level and the second number described Bloom's Taxonomy level for each cell.

The Cognitive Rigor Matrix encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities (Hess, Jones, Carlock, & Walkup, 2009). Webb describes his depth-of- knowledge levels as "nominative" rather than as a taxonomy, meaning that DOK levels name (or describe) four different and deeper ways a student might interact with content (Hess et al., 2009). Moreover, Hess' CRM contains explicit examples of performances tasks and activities students are asked to do in each of the cells matrix which further enhances the reliability of the content analysis. Because Hess' Cognitive Rigor Matrix is designed as a grid, with Webb's Depth of Knowledge as the columns and Bloom's Taxonomy as the rows, a specific matrix was assigned to each cell to provide a more accurate and comprehensive coding scheme. The first number in the matrix described Webb's Depth of Knowledge level and the second number described Bloom's Taxonomy level for each cell. An example of Hess' Cognitive Rigor Matrix is provided in Appendix B, including the following categories and explanations (adapted from Hess, 2009b). See Appendix A for the coding schema breakdown by Webb's and Bloom's categories.

Training and Calibration

To further validate the coding procedures, the two coders concurred that any question that had language that associated with Categories 3 and 4 of Webb's levels would be considered higher level; adhering to the guidelines of the Webb Alignment Tool training manual (Webb, et al., 2005). Similarly, any questions that had language that related to Categories 1 and 2 of Webb's levels, according to Hess' Cognitive Rigor Matrix, would be deemed lower level. Both coders were trained through calibration exercises, that were guided by an experienced coder during the first two coding sessions. During these calibration sessions, the two coders were trained in deductive coding.

In this study, the two coders analyzed each cell of the Cognitive Rigor Matrix prior to coding individually in order to provide further validity and clarity on the types of questions and tasks that would be placed into each category from the PARCC 10th Grade English Language Arts Practice test and Geometry PARCC Practice Test. In order to increase validity and reliability throughout the data collection process, the coders explained their selections in their cell placement for each question. This strategy also aided in the calibration sessions, in which the two coders discussed how each question aligned with specific criteria.

Moreover, some sample rules were adapted from the Webb's Alignment Training (WAT) Manual that the coders followed when assigning Hess' level of complexity. The rules were as follows:

- One: Restate and summarize results and interpretation of what the reviewers have agreed on and what they have disagreed on.
- Two: If there is a difference in interpretation in the cell level, then the reviewer with experience in teaching that grade level with these standards to further understand how the state's teachers might be interpreting the objective.
- Three: If the viewpoints on the DOK/ Blooms Taxonomy levels of an objective are not in consensus, then the WAT will be deferred to. If coders are not in agreement than they both coders will go with the higher level.

Reliability and Validity

Validity in qualitative research indicates consistency and trustworthiness regarding activities and events associated with the phenomenon as signified by the study results explored in the research (Golafshani, 2003). To assure credibility and internal validity this study employed the double-rater read behind method.

To increase reliability of coding, a main coder and second coder compared results when aligning each question to Hess' Cognitive Rigor Matrix following the double-rater read-behind consensus model. The double-rater read behind consensus model proved effective in coding standards (AASA Journal of Scholarship and Practice, 2015, p. 16).

To establish transferability (external validity) this study employed thick descriptions strategies and relied on exemplars and descriptions from the Webb Alignment Tool and the language and descriptions and examples found on the Hess' Cognitive Rigor Matrix. Dependability was established through triangulation. Confirmability was established through reflexivity or intra- or inter-coder reliability, where applicable (Simon & Goes, n.d.). Below is a description of the strategies utilized to further validity of this study:

Pontoratto describes thick description as helping to refer to "the researcher's task of both describing and interpreting observed social action (or behavior) within its particular context," (Ponterotto, 2006, p. 543).

Member checking increased trustworthiness of results. Member checking, also known as participant or respondent validation, is a technique for exploring the credibility of results. Data or results are returned to participants to check for accuracy and resonance with their experiences (Birt, Scott, & Cavers, 2016).

Reflexivity and inter coder reliability also were used. Intercoder reliability is the widely used term for the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion. Debriefing occurred as part of the member checks and double-rater read behind method. Additionally, the coders used the WAT training manual DOK level descriptors as reference points and exemplars (Webb et al., 2005).

- **Reading Level 1:** Simple skills or abilities are used, including recitation. Basic comprehension and understanding are assessed in the form of paraphrasing and repeating specific details from the text.
- **Reading Level 2:** Some mental processing is used that goes beyond recollection and reproduction of a response. Comprehension and subsequent processing of a text is

assessed, including inter-sentence analysis of inference. At this level students are generally asked to summarize, classify, compare and determine.

- **Reading Level 3:** Deeper knowledge is assessed by students being encouraged to think beyond the text and to explain, generalize, or connect ideas. Students are also required to demonstrate understanding, but more abstract themes such as engaging prior knowledge, reasoning, planning, and constructing an inference are also involved.
- **Reading Level 4:** The central focus is higher level thinking, which can be identified by extension activities, which oftentimes require time outside of the classroom. Students utilizing this level will take information from the text and apply it to a new task, develop a hypothesis, and perform complex analyses of connections.

The Web Alignment Tool denotes the following for the DOK levels in Mathematics:

Level 1 *(Recall)*: "Includes the recall of information such as a fact, definition, term or a simple procedure, as well as performing a simple algorithm or applying a formula. An assessment item would require students to demonstrate a rote response (p. 45)."

Level 2 *(Skill/Concept)*: "Includes the engagement of some mental processing beyond a habitual response. An assessment response would require students to make some decisions as to how to approach the problem or activity (p. 45)."

Level 3 *(Strategic Thinking)*: "Requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. Expectations at this level would include drawing conclusions; citing evidence and developing a logical argument for concepts; explain phenomena in terms of concepts; and deciding which concepts to apply in order to solve a complex problem (p. 46)."

> **Level 4** *(Extended Thinking)*: "Requires complex reasoning, planning, developing, and thinking, most likely over an extended period of time. Level 4 activities include designing and conducting experiments and projects; developing and providing conjectures, making connections between a finding and related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs (p. 46)."

Currently training manual exists for Bloom's Taxonomy or the Revised Bloom's Taxonomy, Anderson and Krathwohl (as cited in Hess, Jones, Carlock, & Walkup 2009), provided action verbs that correspond with each of the six Revised Taxonomy levels (as cited in Hess et al., 2009a). Many of these actions can be directly linked to reading and writing practices through ways in which students exhibit their understanding of the material, predominantly in the form of performance tasks.

The following action verbs provided a basis for placement into each of the Bloom's Revised Taxonomy levels. At level 1 in Bloom's Taxonomy learners will typically construct knowledge from long-term memory, recall, locate and identify. At level 2 of Bloom's Taxonomy, learners will begin to further contrast and contextualize that knowledge. Learners at this level might be asked to clarify, paraphrase or illustrate. At level 3 of Bloom's Taxonomy, learners will carry out procedures or apply knowledge to new tasks. At level 4 of Bloom's Taxonomy, learners will begin to distinguish and differentiate. At this level, learners analyze their learning. At level 5 of Bloom's Taxonomy, learners will evaluate and critique. At level 6 of Bloom's Taxonomy, learners combine elements into a new whole. They will construct new meaning and reorganize elements into new.

In this study, the two coders discussed each category within the Cognitive Rigor Matrix prior to beginning coding to further distinguish the questions and activities found within each area. This discussion aided in being further aligned during the coding process. The outcome of the discussion led both coders to be more objective with the expectations of the progression of the cognitive rigor on the matrix; from lower-level thinking to higher-level thinking.

Data Analysis Procedure

The coding committee met on December 20, 2020 in order to discuss and calibrate to the categories found in Hess' Cognitive Rigor Matrix. The two coders evaluated the Hess' Cognitive Rigor Matrix and Webb's Alignment tool. As mentioned previously, the coders agreed on changing the numbers representing Webb's DOK to letters.

Furthermore, they deliberated the complexity of each assessment question. Coders used the sample questions in the Webb's training manual to further calibrate. Questions that the two coders did not both categorize into the same cell were noted to indicate that there was a difference so that an alignment conversation could occur. During the calibration sessions, the two coders discussed each question in which there was a disagreement in placement. After a conversation between both coders, an agreement was reached on the appropriate placement. The two coders also agreed that if both coders could not reach consensus on the placement of a question, they would agree to the higher of the two placements that the coders gave, following recommendations provided by the Webb Alignment Tool. Furthermore, procedures followed in this study modeled those of similar studies in order to provide a consistent methodology in the topic area (Sydoruk, 2019).

Following the discussion, coders then used the PARCC Languages Arts Practice Assessment Grade 10 as part of their training and calibration. The two coders completed 97 questions with 100% agreement due to discussion of each question during the training. Utilizing the double-rater read behind method, the coding committee evaluated the test questions. The two coders used a coding table (see Appendix D) to further guide the organization of the categories that each test question was placed. These procedures helped provide a method of organization, so that the coders could easily check the alignment between them as part of the double-rater method. Figure 11 illustrates an example of the coding table used for this study. The completed template can be located in Appendix D.

Figure 11

Abridged coding template

PARCC Grade 10 English Language Arts/ Literacy:

Ques.	A,1	A,2	A,3	A,4	A,6	B,2	B,3	B,4	B,6	C,2	C,4	C,5	C,6	D,2	D,3	D,4	D,5	D,6
1.																		
2.																		
3.																		
4.																		
5																		
6																		
7																		
8																		

Note. Template for PARCC practice assessments and Hess' Cognitive Rigor Matrix. Full version is located in Appendix D.

Following the first coding session, the two coders completed 77 questions independently. A second coding session was held on December 4th, 2020 to evaluate questions in sets of 10 and to discuss any disagreements found with the codes. Furthermore, if there were any disagreement the coders used the double rater read behind method to mediate any disagreement. One coder presents their findings to the other coding, in alignment with the double-rater- read- behind consensus model. The second coder then agrees or disputes the finding. In their response they must refer to the Hess CRM and the Webb's Alignment Tool to support their rationale. If the coders could not come to an agreement, they followed the suggestion of Webb's et al (2005), assigning a higher depth of knowledge level in instances where there was not agreement.

The data collected in this process were assessed in accordance to their frequency and distribution. The total number were evaluated to calculate percentage. The coding committee reviewed the PARCC Language Arts Practice Assessment Grade 10 and Geometry, completing 97 questions with 90% exact agreement and 100% consensus by the end of the second session. Out of 6 sets 4 sets were completed with 100% agreement. Coders discussed the alignment to the CRM and commonalities between these types of questions within these sets.

In the Language Arts second set one questions was moved from an [A,2] to [B,2]. We went with the higher coding. In this question the question asks what structural choice contributed to the most suspense. This implicated a deeper understanding and application of the text. In the fourth set of English Language Arts, we had 70% agreement with 100% consensus. The second question in this set was moved from and [A,1] to an [A,2]. The question asked for the meaning of the word in context. In the seventh question in this set the question was moved to the higher category from a [B,3] to [B,4]. In the eighth question in this set a question was moved from [C,2] to [C,4]. This was a writing prompt asking students to consider the two essays presented within their analysis. There was an indicator of some ambiguity within this question set, as such we went with the higher code.

In the Geometry question one in the third set of questions was moved from an [A,2] to an [A,3]. This question asks to approximate measures, as referenced by a graph table. In question five of this set the question was moved to a higher code. The question was moved from [B,2] to [B,3].

A third calibration session was conduction December 5, 2020. During this session the coders reviewed the remaining English Language Arts questions. A group of 9 questions were evaluated with 100% agreement and 100% consensus.

After all the data were coded, the frequency and distribution were analyzed to calculate a percentage. This percentage would indicate the percentage of cells of (DOK/Bloom's Taxonomy Levels) as similar to Burn's (2017) formula.

Chapter Summary and Subsequent Chapter

Chapter III described the coding protocol used to align 97 English language arts and Geometry questions (including the divided parts of some questions) from the PARCC practice assessment to Hess' Cognitive Rigor Matrix. For this study, the primary method used was qualitative content analysis methodology used to code each of the PARCC assessment questions.

Furthermore, quantitative descriptive statistics were then used to explain the differences and similarities in the type of thinking associated with the language of the PARCC practice assessment.

Chapter IV presents the findings of this study focusing on the overarching question and the three sub-questions.

Chapter IV: Results

The following chapter presents the findings of the study on the type of thinking that is described in the 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grade 10 Language Arts and Geometry. This study aimed to identify and define the percentage of higher order thinking, as described by Hess' Cognitive Rigor Matrix, embedded in the Partnership for the Assessment of Readiness for College and Career (PARCC) high school practice tests for Grade 10 English Language Arts and Geometry. The coding committee held four coding sessions that took place between December 2nd, 2020 and December 5th, 2020. During these sessions the two coders employed Hess' Cognitive Rigor Matrix as an alignment tool to help categorize and distinguish the higher order thinking that was evident in the PARCC Practice assessments. In the matrix, high order categories include [C,2], [C,3], [C,4], [C,5], [C,6], [D,2], [D,3], [D,4], [D,5], and [D,6].

Two coders analyzed the assessment questions utilizing the double- rater read-behind consensus model. This ensured that the generated results met the accepted criteria defining reliability, by quantitatively defining the degree of agreement between the two coders.

Analysts coded the questions, reviewed the outcomes, and noted agreement and disagreements for each assessment in each grade level. The Web Alignment Tool (WAT) was employed as a secondary support in measuring alignment in Hess' CRM and the PARCC Practice Tests in Language Arts and Geometry and to increase reliability. Any disagreements were later discussed in respect to the criteria and a consensus was then reached. The results of the coding sessions were then calculated to explain the levels of higher order thinking that were found in the language of the assessment.

Deductive content analysis was an analytical method that was employed to test existing categories, concepts, models, theories or hypotheses in the evaluation of cognitive complexity found within the PARCC assessments.

Deductive content analysis is beneficial for testing concepts, categories, theories or any conceptual structure in a new context. Deductive content analysis is usually applied when there is prior theoretical knowledge as a starting point. As such, the research questions are influenced by prior knowledge, and hence, affect the data collection stage. Deductive content analysis is guided by a half-structured or structured analysis matrix (Kyngäs & Kaakinen, 2020). In this case, Hess' Rigor of Matrix was used to structure this form of content analysis. The Hess' Rigor of Matrix uses a cross section of both Bloom's Taxonomy and Webb's DOK that developed a framework for the categorization of higher order thinking. The PARCC practice assessments employed a categorization matrix that represented the concepts found that aligned to higher order thinking. The PARCC practice assessments could then be coded to determine its level of cognitive complexity. The team coded 97 total questions; 58 language arts questions and 39 geometry questions.

The study was guided by the overarching question: What types of thinking are assessed by the questions on the 2019 PARCC practice tests in English language arts and Geometry in grade 10? Hess' Cognitive Rigor matric was employed to assess the thinking requirements of each question in both language arts and geometry. There were three sub questions that further disseminated the guiding question into qualitative and quantitative findings.

Language Arts

The first sub question was: *In what way(s) does the language of the questions on the English language arts section of 2019 Partnership for Assessment of Readiness for College and*

Careers (PARCC) practice tests in Grades 10 associate with the language that promotes higher-order thinking found in research literature?

The lowest level of cognitive rigor that is found on the Cognitive Rigor Matrix is placed at Level 1. The expectation at this level is basic recalling, recognizing, and/or locating basic facts, terms, details, events, etc. Tasks do not require extended thinking and generally require little responses beyond basic recall. Copying, computing, defining, and recognizing are also typical Level 1 tasks (Roach, Elliot, & Webb, 218-231). The complexity of the tasks itself increases as one moves down the matrix within the Bloom's Taxonomy levels. In the lowest level of cognitive complexity, Hess' Cognitive Rigor Matrix contains cells [A,1], [A,2], [A,3], [A,4] and [A,6].

Three questions from a total of 58 on the assessment were placed in [A,1] (Webb's Depth of Knowledge Level 1, Bloom's Taxonomy Recall) cell of the matrix representing 5.17% of the language arts questions. The questions consisted of locating basic facts. An example of an [A,1] question found on the assessment was the following: "What is the meaning of the word arrogates as it is used in the sentence?" The answer was directly evident in the text and deals with identification. Students were asked to select the correct answer consisting facts from the text labeled from A-D.

A total of 22 questions were placed in [A,2](Webb's Depth of Knowledge Level 2, Blooms Taxonomy Understand/Literal Comprehension), accounting for 37.93% of the total number of language arts questions examined. One example of this type of question was "Which quotation from paragraph 3 helps clarify the meaning of resonant?" This question provided the location of the quotation. It expected students to select appropriate words when intended meaning/definition is clearly evident. [A,2] was the modal response for the language arts questions. Consequently, no other questions fell in the Level 1 category or corresponding [A,3] or [A,4] matrix.

Level 2 of Hess' Cognitive Rigor Matrix, aligns with Webb's Level 2 Depth of Knowledge. Webb's Level 2, "includes the engagement of some mental processing beyond a habitual response This level generally requires students to contrast or compare people, places, events and concepts; convert information from one form to another; classify or sort items into meaningful categories; describe or explain issues and problems, patterns, cause and effect, significance or impact, relationships, points of view or processes (www.MDE.K12.MS.us, 2009, p. 9). The questions placed in the Level 2 category required some mental processing beyond recalling. The Cognitive Rigor Matrix contains the following four cells representing the second level of cognitive complexity [B,2], [B,3], [B,4], and [B,6].

Of the 58 questions on the language arts exam, 19 were in the [B,2] category which accounted for 32.76% of the questions on the exam. [B,2**]** corresponds to Webb's Depth of Knowledge Level 2, Bloom's Taxonomy Understand/Literal Comprehension. The [B,2] category was the second highest category found in terms of frequency found within the language arts exam. Many of the questions found in the criteria for [B,2] askes students to: specify, explain, show relationships; explain why (e.g., cause-effect), give non-examples/examples, summarize results, concepts, ideas, make basic inferences or logical predictions from data or texts, identify main ideas or accurate generalizations of texts, locate information to support explicit-implicit central ideas (Hess, Jones, Carlock, & Walkup, 2009)**.** One example of one such question was, "How are the details about Mie's interest in the red cranes important to the development of a central theme in the passage?" In this type of question students are asked to make basic inferences about details as they related to a character's interest. In this case and in all of the other cases in the [B,2] category students are not asked to do further mental processing beyond literal comprehension.

A total of eight questions, or 13.79% of the language arts questions, were categorized into cell [B,3] (Webb's Depth of Knowledge Level 2, Bloom's Taxonomy Application). Mental processes needed for the [B,3] category asks students to use context to identify the meaning of words/phrases, obtain and interpret information using text features, develop a text that may be limited to one paragraph, or apply simple organizational structures (paragraph, sentence types) in writing. An example of a question that aligns to [B,3] was the following: "How does the reference to Sparta in paragraph 11 help to advance the argument of the majority opinion as a whole?" In this example students are asked to interpret information using text features that further develop the text.

Only two questions were categorized in the [B,4] (Webb's Depth of Knowledge Level 2, Bloom's Taxonomy Analysis) category which accounted for 3.45% of the language arts questions analyzed. Mental processes in this category ask students to categorize/compare literary elements, terms, facts/details, events, identify use of literary devices, analyze format, organization, & internal text structure (signal words, transitions, semantic cues) of different texts, distinguish: relevant-irrelevant information; fact/opinion, identify characteristic text features; distinguish between texts, genres. An example of this type of question found within the 10th grade PARCC Language Arts Practice tests is "Refer to the passage from Tinker v. Des Moines Independent Community School District by Justice Abe Fortas and the transcript from Supreme Court Landmark Series: Tinker v. Des Moines. Then answer question 19.- Which paragraph from the majority opinion written by Justice Abe Fortas makes a point similar to the answer to Part A?" In this question students are being asks to distinguish between two texts and compare elements of both cases. Correspondingly no questions analyzed in this language arts practice test found any questions that fell in the [B, 6] category.

In level 3 students are requires to use higher level thinking processes including analysis and evaluation to engage with real world problems that have predictable outcomes. Students are required to provide evidence in reasoning at this level. The cells in Level 3 include [C,2], [C,3], [C,4], [C,5], and [C,6]. At this level of Hess' Cognitive Rigor Matrix, questions were only placed into cells [C,2], [C,3] and [C,4].

There were two questions categorized in [C,2] (Webb's Depth of Knowledge Level 3, Bloom's Taxonomy Understand) cell of the CRM, which accounts for 3.45% of the questions analyzed. An example of a question that aligned to [C,2] included the following: "Both the passage from Woman on the Other Shore and the passage from "A White Heron" explore the central idea of a child's solitude. How is this idea developed differently in the two passages?" Students in this question are required to connect ideas between the two passages and identify inferenced about the overall idea. This question was similar to other questions found within in this cell. Questions placed in this category were asked to support their ideas, make inferences and give examples.

Additionally, one question was found in each of the [C,3] and [C,4] categories accounting for 1.72% of all questions analyzed. The [C,3] (Webb's Depth of Knowledge Level 3, Bloom's Taxonomy Apply) cells, ask students to apply a concept in a new context, revise final draft for meaning or progression of ideas, apply internal consistency of text organization and structure to composing a full composition, apply word choice, point of view, style to impact readers' /viewers' interpretation of a text. One example of this type of question is, "After discovering that his wife has gone missing from the bicycle they were sharing, Mr. Harris returns "to where the road broke into four" and seems unable to remember where he has come from. Using what you know about Mr. Harris, write a narrative story that describes how he chooses which road to take and the experiences he has on his return journey. Be sure to use details from the passage in developing your narrative." Students are asked to apply internal narrative consistency of the text in a developed narrative.

Students must apply their knowledge of the author's craft to maintain consistency and further develop the narrative. The [C,4] (Webb's Depth of Knowledge Level 3, Bloom's Taxonomy Analyze) cells, which consisted of 1.72% of all questions analyzed, asks students to analyze information within data sets or texts, analyze interrelationships among concepts, issues, problems, analyze or interpret author's craft (literary devices, viewpoint, or potential bias) to create or critique a text, and/or use reasoning, planning, and evidence to support inferences. One example of this question is, "Write an essay analyzing the arguments of those who believe certain kinds of speech should be prohibited within an educational setting and those who believe the opposite. Base the analysis on the specifics of the Tinker v. Des Moines case and the arguments and principles set forth in the sources. The essay should consider at least two of the sources presented." In this question type students must be able to interpret the arguments set forth within the two cases and then come to their own interpretation using reasoning and evidence to support their claims. Half of the questions found in the Level 3 area were apparent in writing tasks.

The highest level of cognitive complexity in Hess' Cognitive Rigor Matrix is found within Level 4. At this level, students are required to use reasoning and planning to explain their conclusions. At this level, students also will participate in activities over an extended period of time (Webb, 2007). The coding committee discussed that Level 4 would be difficult to achieve in an assessment with limited time that mostly include multiple choice questions. The cells in Hess' Cognitive Rigor Matrix for Level 4 included [D,2], [D,3], [D,4], [D,5], and [D,6]. Of all questions analyzed in the language arts test, no questions corresponded to these cells.

Geometry

The second sub question was: *In what way(s) does the language of the questions on the Geometry section of 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 10 associate with the language that promotes higher-order thinking found in research literature?*

Hess' Cognitive Rigor Matrix for Math and Science was utilized to assess the thinking requirements in the PARCC Geometry Practice Test. As mentioned previously, the lower level of cognitive complexity contains 5 cells, and accounts for a large number of questions in the Geometry practice tests. Out of 39 questions examined in the Geometry tests 41.03% equated to all of the questions placed into Level 1 cells.

There were three questions placed in the [A,1] cell of the matrix accounting for 7.69% of all of the questions analyzed. Students mental processes in this level should reflect: recall, recognize, or locate basic facts, ideas, principles, recall or identify conversions between representations, numbers, or units of measure, identify facts/details in texts. One example of this question is "In this figure, triangle GHJ is similar to triangle PQR. Based on this information, which ratio represents?" In this type of question, students are expected to use basic recall in or to identify the answer.

There were two questions placed in the [A,2] cell of the matrix accounting for 5.13% of all questions analyzed. Students mental processes in this level are: evaluate an expression, locate points on a grid or number on number line - attend to precision, represent math relationships in words, pictures, or symbols)- attend to precision, read, write, and compare decimals in scientific notation - attend to precision; repeated reasoning. One example of this question was “Find the coordinates of point Q in terms of a, b, and c. Enter your answer in the space provided. Enter only your answer.” In order to answer this question students must locate points on a grid.

There were 11 questions placed in the [A,3] cell of the matrix accounting for 28.21% of all questions analyzed. Students’ mental process in this level are: follow routine multi-step procedures (e.g., long division), calculate, measure, apply a rule (e.g., rounding), apply algorithm or formula (e.g., area, perimeter), solve linear equations, make conversions among representations or numbers, or within and between customary and metric measures- attend to precision. An example of this type of question is “When, 0=28 degrees what is the distance from point A to point B to the nearest tenth of a foot?” Students must apply an algorithm or formula in order to find this answer.

The questions in Level 2 required students to make some decisions as to how to approach a problem. Out of 39 questions examined in the geometry tests, 56.41% equated to all of the questions placed into Level 2 cells. There were 18 questioned placed in the [B,2] cell of the matrix accounting for 46.15% of all questions analyzed. In this category students must: specify and explain relationships, give non-examples/examples, make and record observations, take notes, organize ideas/data, make basic inferences or logical predictions from data or texts Identify main ideas or accurate generalizations. One example of this type of question was: “Which statement about the image of lines AC and PQ would be true under the dilation?” Students here must specify relationships of the lines in order to answer the question.

There were four questions placed in the [B,3] cell of the matrix accounting for 10.26% of all questions analyzed. Students in this category must: select a procedure according to task needed and perform it, solve routine problem applying multiple concepts or decision points, retrieve information from a table, graph, or figure and use it to solve a problem requiring multiple steps, use models to represent concepts, and/or write paragraph using appropriate organization, text structure, and signal words. One example of this type of question was: "Without changing the measurements of the base of the shed, the builder is also considering using a roof angle that will create a roof surface area that is 10% less than the area obtained in Part A. Less surface area will require fewer roofing shingles. Will such an angle meet the specified drainage requirements? Explain how you came to your conclusion. Enter your answer and your explanation in the space provided." In order to answer this question students must retrieve information (table, graph, figure) and use it solve a problem requiring multiple steps.

Level 3 is considered higher-level thinking and require students to demonstrate their knowledge by explanation, reasoning, using evidence to find mathematical solutions. There was one question placed in the [C,2] cell of the matrix accounting for 2.56% of all questions analyzed. In this category students must: explain, generalize, or connect ideas using supporting evidence, explain thinking when more than one response is possible, explain phenomena in terms of concepts, write full composition to meet specific purpose, and/or identify themes One example of this type of question is: "Daniel wants to reshape the other chunk of clay to make a set of clay spheres. He wants each sphere to have a diameter of 4 inches. Find the maximum number of spheres that Daniel can make from the chunk of clay. Show your work." In order to answer this type of question students must connect ideas and use supporting evidence.

Quantitative Findings

The third sub question was: *What is the distribution of thinking on the 2019 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in English language arts and Geometry in Grades 10?*

The expert coders agreed that questions categorized as Level 3 and Level 4 of the Hess; Cognitive Rigor Matrix consists of higher order thinking questions and would be placed as agreed. The cells for Level 3 consisted of C,2], [C,3], [C,4], [C,5] and [C,6] and Level 4 consisted of cells [D,2], [D,3,], [D,4], [D,5] and [D,6]. Out of the 220 questions analyzed 58 were language arts questions and 39 were math questions (see Tables 1 & 2).

Language Arts

Table 1 shows the distribution of thinking requirements on the Language Arts Practice Tests for grade 10. The majority of questions for the Language Arts Practice Test for grade 10 fall between DOK level 1 and DOK level 2 under the understand portion of Bloom's taxonomy. See Figure 9 for overall percentage. 37.93% of questions make up the [A,2] cell category and 32.76% make up the [B,2] cell category. This accounts for 70.69% of the overall distribution of questions asked.

Table 1

Distribution of Thinking Requirements on the Language Arts Practice Tests for grade 10

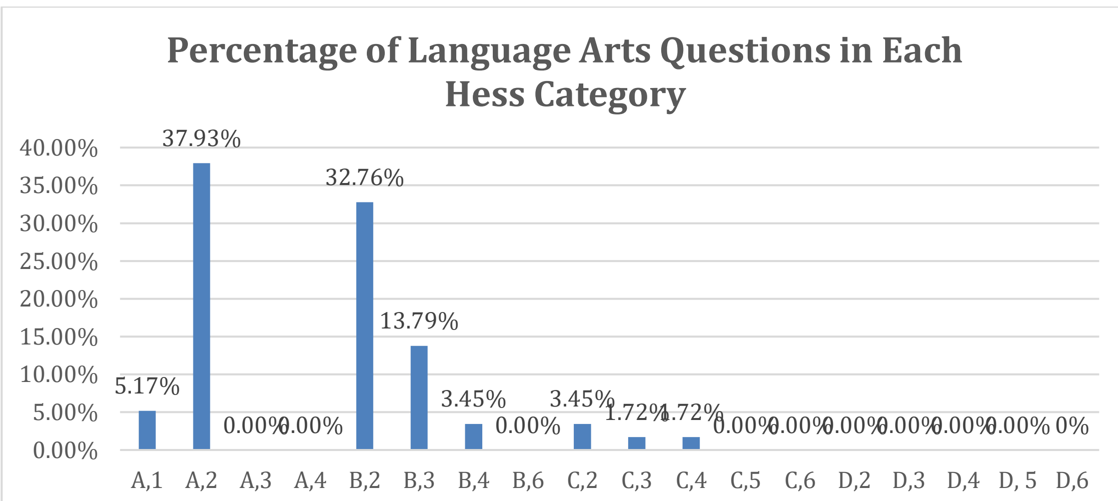
DOK 1	DOK 1	DOK 2	DOK 2	DOK 2	DOK 3	DOK 3	DOK 3
Remember	Understand	Understand	Apply	Analyze	Understand	Apply	Analyze

3	22	19	8	2	2	1	1

From the questions analyzed 54 included languages that aligned with lower-level thinking (see figured 12 & 13) this accounts for 93.10% of the total questions analyzed in the 2019 10th grade language arts practice test.

Figure 12

Percentage of language arts question in each Hess category.



Geometry

Table 2 shows the distribution of thinking requirements on the Geometry Practice Test for grade 10. The majority of questions for the Geometry Test fall between DOK level 1 and DOK level 2 under the understand and apply portion of Bloom's taxonomy. See Figure 10 for overall percentage. 46.15% of questions make up the [B,2] cell category and 28.21% make up the [A,3] cell category. This accounts for a majority of the overall distribution of questions asked on the Geometry test and these two cell categories make up 74.36% of the overall questions asked.

Table 2

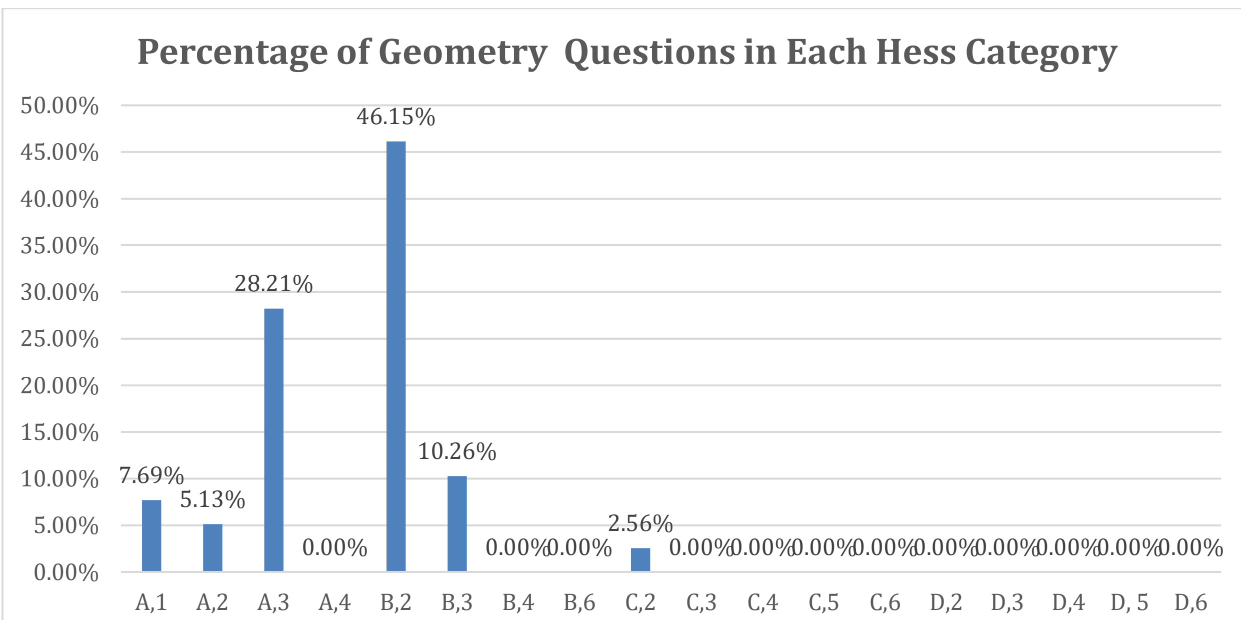
Distribution of Thinking Requirements on the Geometry Practice Tests for grade 10

DOK 1	DOK 1	DOK 1	DOK 2	DOK 2	DOK 3
Remember	Understand	Apply	Understand	Apply	Understand
3	2	11	18	4	1

From the questions analyzed 38 included languages that aligned with lower-level thinking (see figured 12 & 13) this accounts for 97.43% of the total questions analyzed in the 2019 10th Geometry practice test.

Figure 13

Percentage of Geometry question in each Hess category

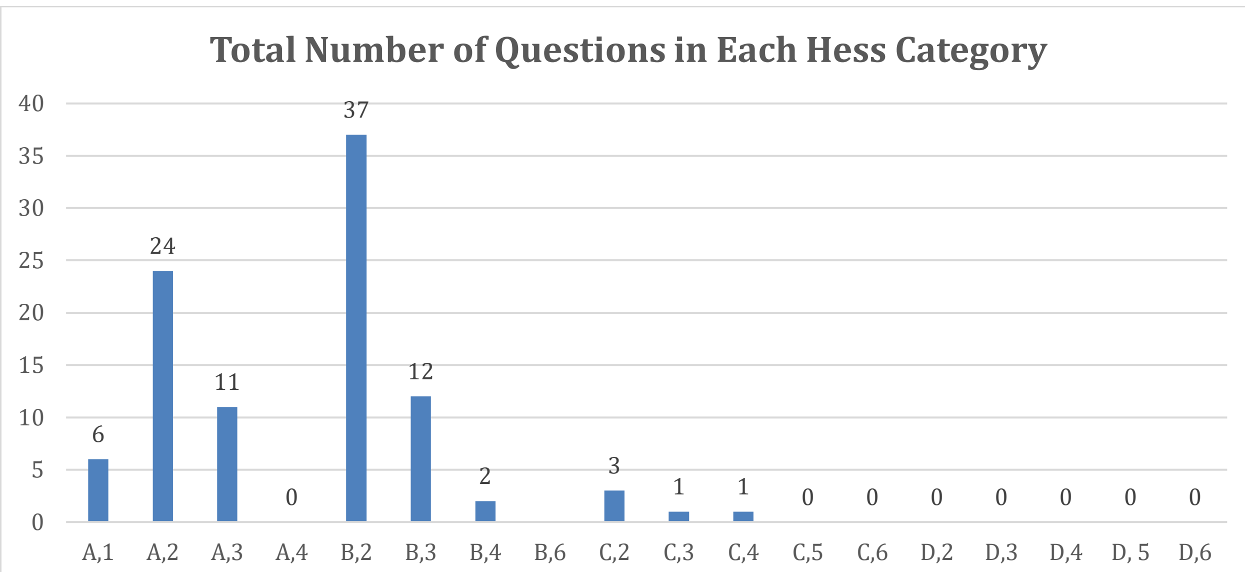


Overall Assessment Findings

The cell with the greatest overall frequency was [B,2] that has 37 questions making up 38.14% of the total questions analyzed. See Table 11. Cell [B,2] represents Webb's DOK level 2 and Bloom's Taxonomy Understand. DOK Level 2 is about forming a conceptual understanding of a topic and generally refers to the integration and application of concepts and other ideas and does not extend any further into elevated areas of higher order thinking.

Figure 14

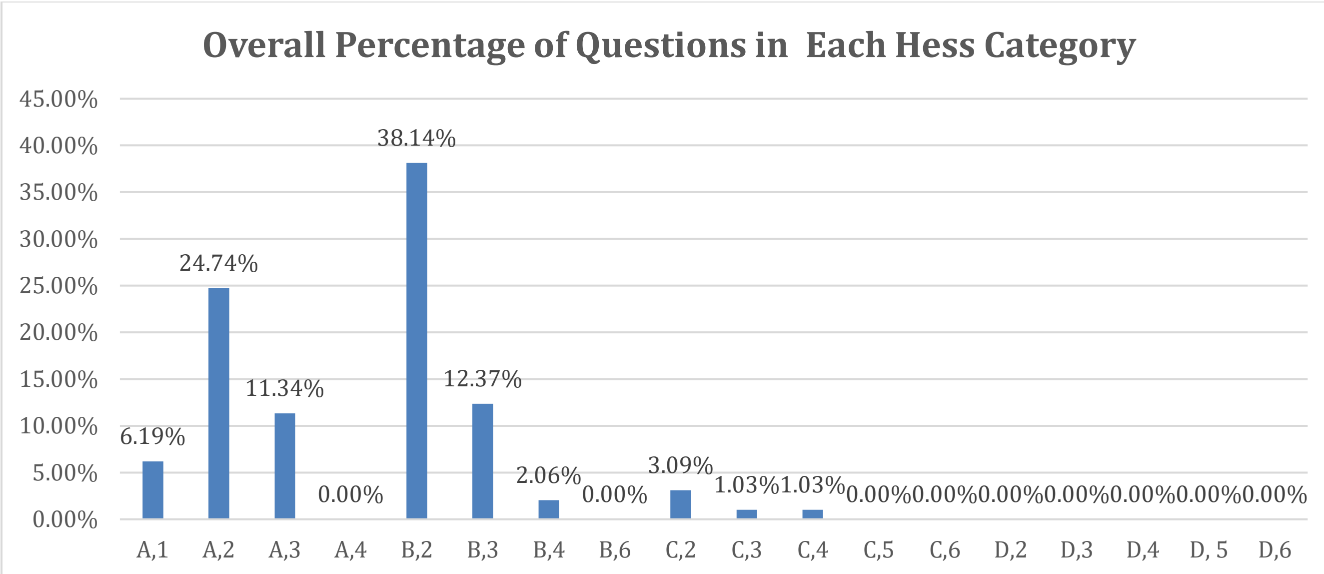
Total Number of questions in each Hess Category



The overall distribution of questions asked between both English Language Arts 10th Grade Practice test and the Geometry test fell predominately in the understand level. These questions accounted for 62.88% of the overall distribution. See Figure 14. The cell with the greatest frequency was [B,2] accounting for 38.14% of the total questions analyzed.

Figure 15

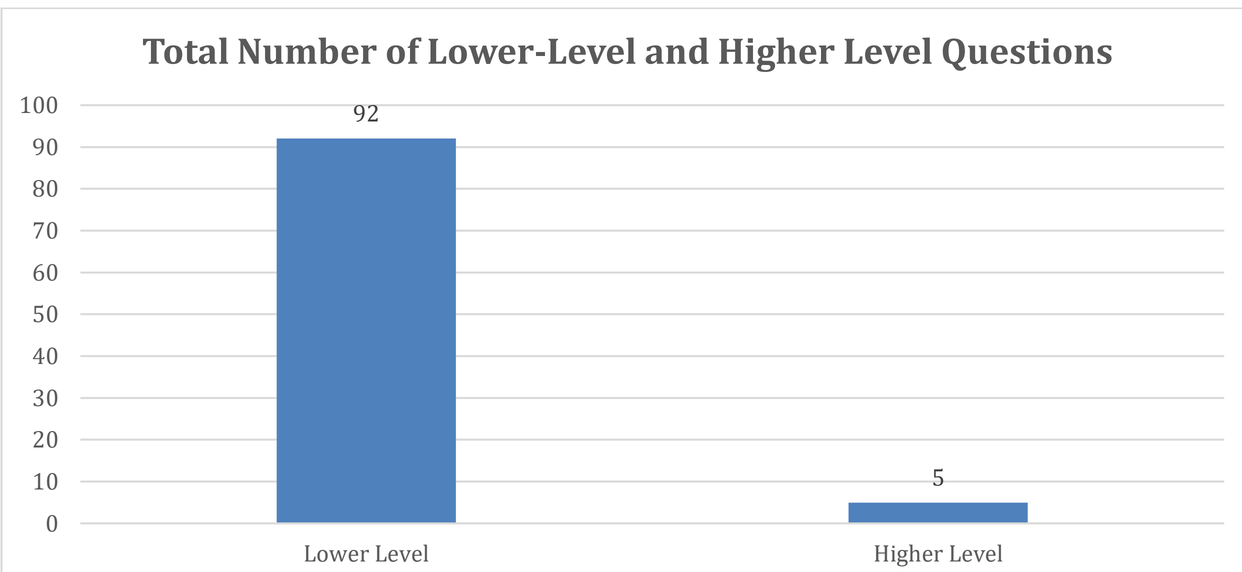
Percentage of Questions in Each Hess Category



Total Assessment

Figure 16

Total number of lower-level and higher-level questions



Questions placed in this category required students to (Hess et al., 2009):

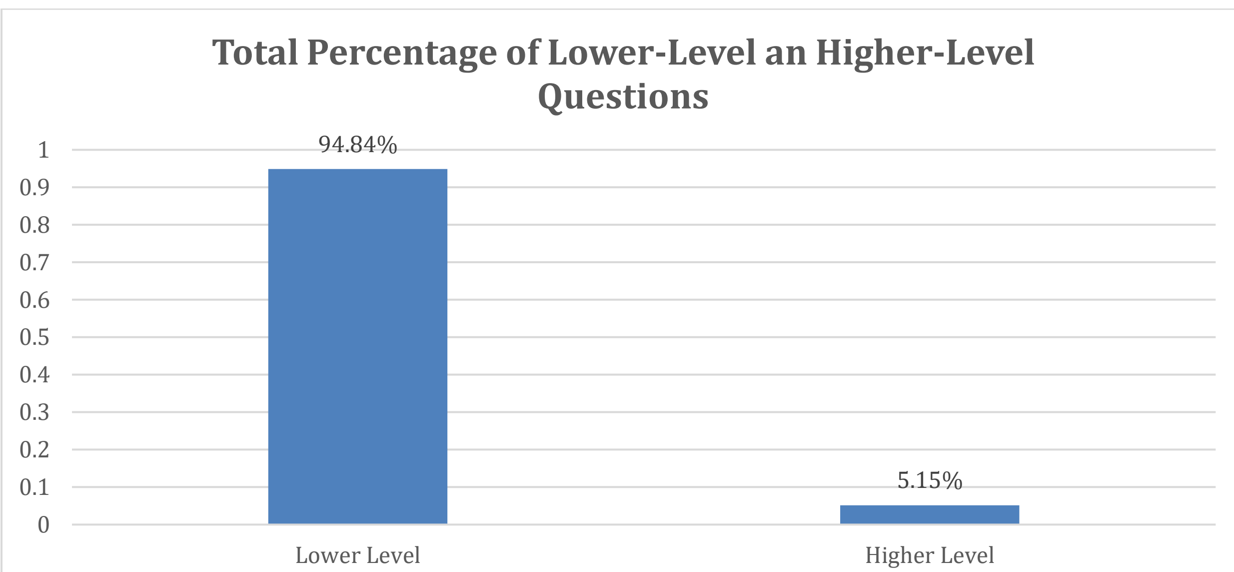
- Specify explain, show relationships (e.g., cause-effect)
- Give examples/ non-examples

- Summarize results, concepts, ideas
- Make basic inferences or logical predictions from data, texts, or observations.
- Identify main ideas or accurate generalizations of texts
- Apply simple organizational structures
- Use models/diagrams to represent of explain mathematical concepts
- Make and explain estimates

[B,2] as mentioned previously, includes the engagement of mental processing, even though it is still considered lower-level thinking. Level 2 questions still require students to demonstrate past a rote response or beyond basic recall like in Level 1.

Figure 17

Total Percentage of Lower-Level and Higher-Level Questions



94.84% of the questions were categorized as lower-level questions requiring students to recall, reproduce, and use skills. 5.15% of the questions analyzed were categorized as cognitive complex requiring strategic thinking, reasoning, and extended thinking.

Conclusion

The purpose of this mixed method was to evaluate the language found in the English Language Arts and Geometry sections of the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in grades 10 and how it compares with the language that associates with higher-order thinking found in research literature. The 10^{th} grade was selected for this study due to the lack of existing research.

In response to the research questions, the data analyzed suggested the following trends from the 97 questions taken from the Language Arts and Geometry PARCC Practice Tests in grades 10:

- Out of 97 questions, 92 were categorized as lower questions equating to 94.85%, of all questions analyzed.
- Out of the 97 questions, 5 were categorized as higher-level questions equating to 5.15% of all questions analyzed.
- The cell with the highest level of frequency was [B,2] which had 37 questions making up 38.14% of the total questions analyzed.
- No questions were placed into Level 4. The most cognitively complex questions in the PARCC practice assessments were placed into cell [C,4].

Chapter V includes a summary of the methodology and a discussion of the findings as they relate to the three sub questions, as well as implications for policy and practice, and future research recommendations.

Chapter V: Conclusion

This chapter provides a summary of the study, including observations on the findings as they connect to the overarching research question and both sub questions, a conclusion, implications for policy and practice at a district level and recommendations for future study. The purpose for this qualitative content analysis study was to describe and compare the extent of cognitive complexity, as defined in Hess Cognitive Rigor Matrix, embedded in the Partnership for the Assessment of Readiness for College and Career (PARCC) practice tests for English Language Arts and Geometry. For this study, a total of 97 questions from both the English Language Arts Grade 10 PARCC Practice Test and Geometry PARCC Practice test, were examined using Hess' Cognitive Rigor Matrix. No empirical evidence exists on the higher order thinking requirements embedded within the PARCC Test. Similarly, there is no evidence on the categorization of the types of questions provided in the PARCC Test in regards to the type of thinking or level of cognitive complexity required of students.

Methodology Summary

Hess' Rigor of Matrix was used as an analytical framework in this study to categorize the language of cognitive complexity found on the PARCC 10th grade ELA and Geometry Practice Tests. Complexity resides not only in the demands of problem representation, but also in the levels of knowledge and application needed to formulate a problem solution

The Cognitive Rigor Matrix (CRM) couples the higher order thinking of Webb's Depth of knowledge and the analysis of cognitive skills within tasks and assessments. The CRM directly correlates and clearly articulates the two schemata. This tool enables educational stakeholders to evaluate instructional tasks and their complexity levels (Hess, et al., 2012). Both Webb's Depth of Knowledge and Bloom's Taxonomy are the two schemata utilized in this matrix. At the lowest level, Level 1, there is very little cognitive complexity. Questions on this level rely on basic operations that do not require analysis of the text (Webb et al, 2005, p. 70). Questions in Level 2, the second level of Webb's Depth of Knowledge are also considered lower in cognitive complexity as this level asks learners to do basic application and understanding that does not extend into analysis or extended thinking. However, Levels 3 and 4 of Webb's Depth of knowledge are considered higher level because they require more complex thinking to answer non routine questions such as reasoning and identifying abstract concepts. Level 4 is considered the highest level of cognitive complexity according to Webb's Depth of Knowledge. At this level, students extend their thinking and learning beyond the question or task being asked. These tasks typically take longer time periods for the students to respond to.

The levels of Bloom's Taxonomy embedded into Hess' Cognitive Rigor Matrix include, from lowest to highest level: remember, understand, apply, analyze, evaluate, create. These levels increase in complexity based on the type of tasks that students are asked to undertake. Tasks with lower cognitive complexity in Webb's Depth of Knowledge could place into a higher level of complexity of the task in Bloom's Taxonomy.

The Cognitive Rigor Matrix encapsulates the complexity of the subject as well as the cognitive engagement with that content (Hess et al., 2009). Webb DOK describes four different and deeper ways a learner might engage with content (Hess et al., 2009). Moreover, Hess' CRM contains explicit examples of performances tasks and activities students are asked to do in each of the cells matrix which further enhances the reliability of the content analysis.

This study used qualitative content analysis methods to describe and categorize distribution of replicable and valid inferences using high order thinking in (PARCC) high school practice tests for English Language Arts and Mathematics, based on Hess' Cognitive Rigor

Matrix. The purpose of which was to answer the overarching question: *What are the types of thinking are assessed by the questions on 2019 PARCC practice tests in English language arts and Geometry in grades 10?*

Two coders used deductive category application, in accordance with Mayring's (2000) Step Model, to organize each question into the appropriate cell of Hess' Cognitive Rigor Matrix. Each cell was assigned a matrix based on the level of Webb's Depth of Knowledge and Bloom's Taxonomy. An example of a matrix is [2,4], which is Webb's Level 2 and Bloom's analyze.

Summary and Discussion Findings

As discussed in Chapter 1, the PARCC test claims to emphasize a marked increase in critical analytical thinking and higher order thinking skills. The Common Core State Standards, and the Next Generation Science Standards, emphasizes the need to assess critical analytical thinking (CAT) and to be able to develop a transfer of skills across subject matters. For Grade 11 High School English Language Arts, for instance, the PARCC consortium claims that, "The PARCC complexity framework reflects the importance of text complexity as it relates to the CCSS, (Common Core Curriculum Standards), which indicates that 50 percent of an item's complexity is linked to the complexity of the text(s) used as the stimulus for that item… To this end, PARCC has developed a clear and consistent model to define text complexity and has determined to use three text complexity levels: readily accessible, moderately complex, or very complex" (PARCC, ETS, PEARSON, Assessment SIG Business Meeting , 2014). However, the "consistent model" is PARCC's own interpretation of complexity and does not align to the empirical literature.

Of the 97 questions examined in both the 10th Grade English Language Arts PARCC Practice Exam and the Geometry PARCC Practice Exam 94.84% of the questions were categorized as lower-level questions requiring students to recall, reproduce, and use skills, and/ or concepts and 5.15% of the questions analyzed were categorized as cognitive complex requiring strategic thinking, reasoning, and extended thinking.

From the questions analyzed from the 54 included languages that aligned with lower-level thinking (see figured 9 & 10) this accounts for 93.10% of the total questions analyzed in the 2019 10th grade language arts practice test. From the questions analyzed 38 included languages that aligned with lower-level thinking (see figured 9 & 10) this accounts for 97.43% of the total questions analyzed in the 2019 10th Geometry practice test.

Conclusion

The PARCC test claims to emphasize a marked increase in critical analytical thinking and higher order thinking skills. The Common Core State Standards, and the Next Generation Science Standards, emphasizes the need to assess CAT and to be able to develop a transfer of skills across subject matters. Upon analyzing the questions from the test, the PARCC consortia claims are not valid. The finding suggest that the assessment quests are not cognitively complex and provide a widely lower level of cognitive complexity. This study provided insight into what assessments illicit within a student's thinking demands.

It further empathizes that current assessments often do not develop appropriate frameworks to establish a valid measure that assesses higher order thinking skills. Moreover, measurement strategies such as, subjective rubrics and item constructions, do not often provide the appropriate tools to reflect a learner's true progress (Development Process, 2018). A principled approach to assessment design is critical to ensure accuracy, as well as cognitive rigor that mirrors a student's ability to transfer skills. Educational stakeholders have been long believed in the benefits of these standardized assessments but have failed to further validate and investigate whether or not the claims of cognitive complexity within these assessments are true. The limitations of the PARCC tests start with the overall multiple-choice format, which leads students to the correct answer or does not illicit an extended thinking task that is higher level in nature. Similarly, the types of questions that were asked of students routinely fell within the [B,2] category that is considered lower level. This level category consisted of 37 questions making up 38.14% of the total questions analyzed.

It is important that the questions on standardized assessments are inspected at a state and local level to determine if the items match the claims. This process allows to further vet testing according to the instruction provided, curriculum implemented, embedded bias, as well as the appropriate variance of cognitive complexity. Tienken (2011) emphasized that "it is dangerously naïve and professionally irresponsible to think that one set of standards, based solely on two subjects, can prepare children to access the thousands of college options or even make them attractive to the admissions officers that control access to those options" (p. 11).

High-stakes testing regimes like that of PARCC and NJSLA influence schools and education at all levels. Their impact on teaching practices, distribution of resources and curriculum provision, and whether they achieve the intended increases in student achievement in targeted areas. Modification of teaching and curricular practices in response to test preparation often supersedes best practices for student learning and locally adapted curriculum. The findings of this study further confirm the capacity of high-stakes regimes to distort their effectiveness in assessing students' higher order thinking capacity.

While students engage in higher-order thinking as they read complex texts and perform complex reading- related tasks, even the most consequential assessments, high-stakes tests, are currently limited in providing information about students' higher- order thinking (Afflerbach et al., 2015). Moreover, the complexity of intellectual engagement with the text is not held within the text itself but, instead, in the demands placed on the reader by the teacher's questions. (Degener & Berne, 2016). Meaning that contextualization and personalized learning that is derived from the teacher student exchange is what further moves student outcomes and further increases higher order thinking not teaching to the test.

Too often test-centered curricula are often heralded as an effort to decrease achievement gaps, although analyses have demonstrated that the effort has largely failed (Braun et al., 2014) and has led instead to academic disengagement (Moon et al., 2007). Conversely, it is actually higher order thinking and problem-solving skills that have been positively correlated with increased test scores (Wasserberg & Rottman, 2016). Therefore, if 94.84% of the questions assessed in this study were categorized as lower-level questions then not only do PARCC tests further lead to disengagement they also fail to increase higher order thinking. The PARCC questions do not allow students to move to an extended way of thinking as evidenced in DOK levels 3 and 4.

Recommendations for Leadership Practice

The findings of this study revealed that the PARCC assessments must be revised to better gauge students in connection with high level cognitive complexity. Moreover, educational stakeholders must look at the impact that PARCC tests have on policies and curricular practices to revisit a more holistic approach to better serve students and their academic outcomes. If standardization does persist, then at the very least it should include both lower-level and higher-order questions to provide a more in depth understanding of a student's capabilities at their given grade level. Once more, over emphasizing the cognitive domain in the hopes that it will improve testing outcomes can be detrimental to overall student outcomes and the at-risk populations within a school. As such, school leaders should also consider instructional opportunities that expands the students' creativity and critical thinking through multiple approaches.

Moreover, school leaders and leadership should not only use matrix's like that of Hess' CRM but they should also approach curriculum design, development and implementation that is guided by evidence and experience of how students learn best and the complementary functions of public school (Tienken C. , 2017, p. 113)

Another consideration might be to further adapt a problem-based curriculum. Studies comparing learning outcomes for students taught via project-based learning (PBL) versus traditional instruction show that when implemented well, PBL increases long-term retention of content, helps students perform as well as or better than traditional learners in high-stakes tests, improves problem-solving and collaboration skills, and improves students' attitudes toward learning (Strobel & Berneveld, 2009, p. 55). Learning progressions may inform the design of tasks so that they are more likely to be both suitable and informative at a particular level. Design principles most commonly used in PBL align well with the goals of preparing students for deeper learning, higher-level thinking skills, and intra/interpersonal skills.

School leaders should instill the use of implementation templates to further clarify curricular expectations. Part of improving and implementing a higher quality of curriculum learning and teaching is to also provide effective professional development. Teachers need time and opportunities to work together in collaborative communities to further reflect on their classroom instruction and practice to further student outcomes.

Recommendations for Policy

Over the course of this century, the concept of teaching has shifted from an industrial model teacher replicating a specific set of instructional tasks to a "complex, dynamic, interactive, intellectual activity" (Smylie & Conyers, 1991, p. 13). This shift occurred for many reasons, one of which was the change to a much more diverse student body, as well as changed to the economy. As such, teachers need to approach instruction with constant reflection, evaluation and experimentation. We also now expect teachers to alter curricula on the basis of new knowledge and ways of knowing, to refection the shifting needs of their local student population (Richardson, 1998)**.**

The ultimate goal of the school system according to the theory of performativity should be efficiency of desired standardized outputs, not quality, creativity, or innovation. This theory has been widely adapted in subscribed to in many policies that are evident today. As a result, policy makers preclude processes, inputs and outputs that are not easily monitored or measured. (Tienken C. , 2017, p. 90)

In 2001, the United States Congress passed the No Child Left Behind Act, (NCLB) which instituted education reform nationwide with the hopes of establishing measurable goals to improve outcomes for all learners. Part of this act required states to develop standardized tests to measure student learning in order to receive federal funding.

The National Education Policy Center (NEPC) at the University of Colorado Boulder emphasized this point in response to policy that encourages standardization. There is no evidence that any test score increases represent the broader learning increases that were the true goals of the policy. Goals of the policy involve critical thinking and student preparedness, as well as cultivating lifelong learners (Welner & Mathis, 2015)**.** Additionally, annual tests rolled out through NCLB haven't done very much in closing the education gap which was the main goal of NCLB.

The Every Student Succeeds Act (ESSA, U.S. Department of Education, 2020) was reauthorized and signed by former President Obama on December 10, 2015. According to the U.S. Department of Education (2020), the previous version of the law, the No Child Left Behind (NCLB) Act. However, tests are still federally mandated along with 95% participation with no opt-out clause as was present in the House bill. There are new mandates for the tests to be psychologically profiling "higher order thinking skills" and "strategies to improve students' skills outside the academic subject areas." Experts admit these standards and assessments are very subjective and there is new evidence of lax student data protection (National Education Association, 2020)**.**

Local education decisions traditionally have been the provenance of states and local districts, but Bush led the way for more federal involvement requiring students in grades 3 through 8 and once in high school to take standardized tests for school “accountability” purposes. The tests were only in math and reading, leading schools to focus on those subjects and precluding other subjects such as history, science, physical education and the arts.

Both the combined components of the instructional core and the psychological factors of learning motivation led to a better understanding of the decision-making process that led to disengagement of students and undesirable academic outcomes. Time and time again, research continues to illustrate that the strongest correlation to student outcomes both academically and otherwise, are related to engagement and efficacy, both of which lay within the hands of the teacher (Fortney, 2016). Additionally, higher order thinking is achieved more frequently and in a higher variance when contextualization and personalization occurs to further illicit a more extended response from a student given their background and learning considerations.

In policy consideration it is important to note that “the ultimate assessment system already exists in public school classrooms: the teacher (Tienken, 2019, p.59). Teachers provide real-time feedback to students and families.

Although there is some political and educational consensus agrees on producing youth with strong higher order thinking skills, accomplishing this task is more challenging than just increasing educators’ motivation, requiring more than financial or accountability incentives (Richland & Begolli, 2016). Yong Zhao (2010) described the approach to return public education to the public as “mass localism.” He explains that local control is the key to creativity because each local can address its needs through customized solutions that draw on evidence-based practices and ideas (Tienken C. , 2017, p. 151).

In March 2020, nearly every school in the nation suddenly adapted remote learning in the wake of the national pandemic. At the same time, states waved the requirements for state testing. However, the Biden Administration on February 22, 2021 announced a return to nationally and state testing. This return was rationalized in that there was a need to address the educational inequities that the pandemic has made worse. However, this research, clearly illustrates the deficits in national testing. National testing does not serve as a diagnostic and its curricular influences can hurt educational outcomes. Instead, administrative and federal policies, should increase funding around local professional development and resources needed to aid in optimizing local education.

Recommendations for Future Research

The intent of this mixed-methods study was to describe the way(s) in which the language found in the English Language Arts Grades 10 and Geometry sections of the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests associate with the language that promotes higher-order thinking found in research literature. To date, there is no empirical evidence existing in regards to PARCC test questions and its preparedness for college and career. Similarly, there is no evidence on the categorization of the types of questions provided in these PARCC assessments regarding the type of thinking or level of cognitive complexity required of students. Further research is needed on the complex thinking embedded in standardized testing.

Further research that evaluates the variance of cognitive complexity within local curriculums and instruction should be analyzed within the language arts and mathematics areas at the high school grade levels. Additionally, research in the area of how curriculums might vertically align to naturally progress and build from one level to the next in a way that might best suit learner outcomes would help aid in strengthening curriculum and instructional strategies.

Longitudinal studies on overall outcomes of students in higher education based on testing performance would be impactful in further identifying the effectiveness of testing and standardization efforts. It might further illustrate the lasting impacts of standardization. This type of study would help to further find long-term patterns of learning and curricular designs that have been influenced by standardized high stakes tests**.** Many research studies on standardized tests focus on short-term data alone. That means long-term data may offer patterns or information that have not been collected previously.

The interconnectedness of national testing and curricular practices at the local level has been thoroughly documented. However, not from the perception of the teacher in regards to how their curricular and daily instructional practices are influenced by test preparation. Furthermore, teachers perceptive of the PARCC test as enhancing curricular practices and student learning is needed to understand the correlation between national testing practices and teacher's perception.

Lastly, further studies should examine the impact that standardization has in lower socio-economic environments. While research has indicated that standardization efforts have not been successful in "closing the learning gaps," more information is needed on the impact that overly stressing the cognitive domain has on a learning population that already has a larger volume of students that are classified as high risk; meaning they are at danger of not graduating.

References

Abdelhamid, T. M. (1999, December 13). The Multidimensional Learning Model: A Novel Cognitive Psychology-Based Model for Computer Assisted Instruction in Order to Improve Learning in Medical Students. *Medical Education Online, 4*(1).

Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association*, 103(3), 152–153. doi:https://doi-org.ezproxy.library.berkeley.org/10.3163/1536-5050.103.3.010

Afflerbach, P., Cho, B.-Y., & Kim, J.-y. (2015). Conceptualizing and Assessing Higher-Order Thinking in Reading . *Theory Into Practice*, 203-212.

Armstrong, P. (n.d.). *Bloom's Taxonomy*. Retrieved from Vanderbilt University: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Astuti, N., Bangli, S. S., & Bali, B. (2018). TEACHER'S INSTRUCTIONAL BEHAVIOUR IN INSTRUCTIONAL MANAGEMENT AT ELEMENTARY SCHOOL REVIEWED FROM PIAGET'S COGNITIVE DEVELOPMENT THEORY. *SHS Web of Conferences* .

Austin, S. L. (2009). Hierarchies Of Abilities And Activity Demands In The nd Allen Diagnostic Module 2 Ed.: A Validity Study . *Graduate College of the University of Illinois at Chicago,* .

Bandura, A. (1993). Percieved Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist, 28*(2), 117-148.

Bell, R. C. (2004, February 15). *Cognitive complexity*. Retrieved April 20, 2018, from THE INTERNET ENCYCLOPAEDIA OF PERSONAL CONSTRUCT PSYCHOLOGY: http://www.pcp-net.org/encyclopaedia/cogcom.html

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *Elseview*, 8-14.

Benjamin, R. (2014, March). Two Questions About Critical Thinking Tests in Higher Education. *Change*, 24-31.

Brookhart, S. M. (2010). *How To Assess Higher Order Thinking Skills In Your Classroom.* Alexandria, VA: ASCD.

Brown, E. (2015, July 22). *As states drop out of PARCC's Common Core test, faithful carry on*. Retrieved from Washington Post: https://www.washingtonpost.com/local/education/as-states-drop-out-of-parccs-common-core-test-faithful-carry-on/2015/07/22/f31eeb48-2f1c-11e5-8353-1215475949f4_story.html

Brown, N. J., Afflerbach, P. P., & Croninger, R. G. (2014, October). Assessment of Critical-Analytic Thinking. *Educ Psychol Rev, 26*, 543-560.

Brown, N. J., Afflerback, P. P., & Croninger, R. G. (2014, October). Assessment of Critical-Analytic Thinking. *Educ Psychol Rev* .

Burleson,, B., & Caplan, S. (1998). Cognitive Complexity. *Communication and Personality: Trait Perspectives*, 233-286.

Burns, C. R. (2017). A Comparison of Complex Thinking Required by the Middle School New Jersey Student Learning Standards and Past New Jersey Curriculum Standards . *Seton Hall University Dissertations and Theses* .

Business Simulations and Cognitive Learning:. (2008, July 31). *Simulation Gaming*. doi:10.1177/1046878108321624

Case, R. (n.d.). The unfortuate consequences of Bloom's taxonomy. *The Critical Thinking Consortium*.

Clark, D. (2015, January 5). *Bloom's Taxonomy of Learning Domains*. (L. D. A Big Dog, Producer) Retrieved April 21, 2018, from Big Dog and Little Dog's Learning Juxtaposition: http://www.nwlink.com/~donclark/hrd/bloom.html

Common Core State Standards. (2020). *English Language Arts Standards Anchor Standards College and Career Readiness Anchor Standards for Reading*. Retrieved from Common Core State Standards: http://www.corestandards.org/ELA-Literacy/CCRA/R/

Danzis, D. (2018, October 18). *The Press of Atlantic City*. Retrieved from PARCC gets new name, fewer questions for spring 2019: https://pressofatlanticcity.com/education/parcc-gets-new-name-fewer-questions-for-spring-2019/article_7c344958-37ff-5b8b-84bd-fc67f7f244a0.html#:~:text=Eno%20announced%20that%20instead%20of,to%20administer%2C%20the%20memo%20reads.

Degener, S., & Berne, J. (2016). Complex Questions Promote Complex Thinking. *The Reading Teacher , 70*(5), 595-599.

Department of Education State of New Jersey. (2016, July). *Department of Education*. Retrieved from Historical Context: Overview of New Jersey's Statewide Testing Program: https://www.nj.gov/education/assessment/history.shtml

Development Process. (2018). (2. C. Initiative, Producer) Retrieved April 18, 2018, from Common Core State Standards Initiative : http://www.corestandards.org/about-the-standards/development-process/

Dobosh, M. (2015, June 2). THE IMPACT OF COGNITIVE COMPLEXITY AND SELF-MONITORING ON LEADERSHIP EMERGENCE. *University of Delaware*.

Dougherty Stahl, K. A., & Schweid, J. (2013). BEYOND MARCH MADNESS - Fruitful Practices to Prepare for High-Stakes ELA Tests. *Research Into Practice, 67*(2), 121–125 .

Eisenman Jr., J. G. (1995). An Evaluation of the Higher Order Thinking Skills Program with Fourth and Fifth Grade Students. *A Dissertation Submitted to the Graduate Faculty of The University of Georgia*.

Elementary and Secondary Education Act (ESEA). (n.d.). *Elementary and Secondary Education Act (ESEA)*. Retrieved from Washington Office of Superintendent of Public Instruction: https://www.k12.wa.us/policy-funding/grants-grant-management/every-student-succeeds-act-essa-implementation/elementary-and-secondary-education-act-esea

Elferchichi,, A., Alaya, N. B., & Ahmed, M. B. (2007, September). Towards a multidimensional model for learning objects and learners. *Kassel University Press*.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *Sage*, 1-10.

Fast, D. (2014, November). Surveying Arizona's Third through Fifth Grade Teachers about their Confidence in Teaching the Cognitive Demands of the Common Core State Standards to All Students .

FORTNEY, V. J. (2016, May). AN EXPLORATION OF FACTORS THAT INFLUENCE STUDENT ENGAGEMENT IN SCIENCE . *Southern Connecticut State University* .

Francis, E. (2017). *What Is Cognitive Rigor?* (ASCD, Producer) Retrieved Apri 21, 2018, from Edge A Professional Network Community for Educators: http://edge.ascd.org/blogpost/what-is-cognitive-rigor

Francis, E. M. (2018, July 23). *Maverik Education*. Retrieved from Depth of Knowledge or Extent of Learning? Cognitive Rigor, Complexity, Deeper Teaching and Learning, Depth of Knowledge: https://maverikeducation.com/blog/f/depth-of-knowledge-or-extent-of-learning

Graf, E. A., & Arieli-Attali, M. (2015). Designing and Developing Assessments of Complex Thinking in Mathematics for the Middle Grades . *Theory Into Practice* , 195-202.

GRUBERMAN, R. S. (2005, August). TEACHER CONCEPTUALIZATIONS OF HIGHER-ORDER THINKING: A CASE STUDY .

Herman, J., & Linn, R. (2013). *On the Road to Assessing Deeper Learning: The Status of Smarter Balanced and PARCC Assessment Consortia.* CRESST/University of California. Los Angeles, CA: CRESST/University of California.

Herman, J., Buschang, R. E., & La Torre Matrundola, D. (2014, October). An Explanation of the ELA and Math Cognitive Complexities Frameworks. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.

Hess, K. K., Carlock, D., Jones, B., & Walkup, J. R. (2009). What exactly do “fewer, clearer, and higher standards” really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments.

Hess, K. K., Carlock, D., Jones, B., & Walkup, J. R. (2009). What exactly do “fewer, clearer, and higher standards” really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments. *E D U C A T I O N A L RESEARCH IN A C T I O N*.

Hess, K. K., Jones, B. S., Carlock, B. S., & Walkup, J. R. (2009, March). Cognitive Rigor: Blending the Strengths of Bloom's Taxonomy and Webb's Depth of Knowledge to Enhance Classroom-level Processes.

Hess, K., Abedi , J., Thurlow , M., Hiebert , E., Darling-Hammond , L., Elliott- Shuman , N., & Lynn Goldberg , G. (2012, January 6). Content Specifications for the Summative Assessment of the Common Core State Standards for English Language Arts and

Literacy in History/Social Studies, Science, and Technical Subjects. *SMARTER Balanced Assessment Consortium*.

Hummer, R. A., & Hernandez, E. M. (2015, May 18). The Effect of Educational Attainment on Adult Mortality in the United States*. *Popul Biull*, 1-16. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4435622/

Initiative, 2. C. (n.d.). *Development Process*. Retrieved April 18, 2018, from 2018 Common Core State Standards Initiative: http://www.corestandards.org/about-the-standards/development-process/

Initiative, 2018 Common Core State Standards. (n.d.). *Development Process*. Retrieved April 18, 2018, from 2018 Common Core State Standards Initiative: http://www.corestandards.org/about-the-standards/development-process/

Jochim , A., & McGuinn, P. (2016, Fall). The Politics of the Common Core Assessments. *Education Next*, *16*(4).

Jochim, A., & McGuinn, P. (2016, July 12). *Education Next*. Retrieved from The Politics of the Common Core Assessments: https://www.educationnext.org/the-politics-of-common-core-assessments-parcc-smarter-balanced/

Karin Hess, E. (2006). Linking Formative Assessment to Instructional Decisions: Taking a Closer Look. *Reidy Interactive Lecture Series (RILS).* Nashua, NH.

Kim, S.-E. (2005, May). EFFECTS OF IMPLEMENTING PERFORMANCE ASSESSMENTS ON STUDENT LEARNING: META-ANALYSIS USING HLM . *The Pennsylvania State University The Graduate School Department of Educational and School Psychology and Special Education* .

King, F., Goodson, L., & Rohani , F. (n.d.). Higher Order Thinking Skills. *Educational Services Program* .

Knauff, M., & Wolf, A. G. (2010, March 10). Complex cognition: the science of human reasoning, problem-solving, and decision-making. *Cogn Process, 11*, 99-102.

Kyngäs, H., & Kaakinen, P. (2020). Education Inquiry. *The Application of Content Analysis in Nursing Science Research: Deductive Content Analysis*, 1. doi:10.1080/20004508.2020.1855827

Marzano, R. J., & Kendall, J. S. (2007). *The New Taxonomy of Educational Objectives.* Thousand Oaks, CA: Corwin Press.

Maverick Education. (2018, July 23). *Depth of Knowledge or Extent of Learning?* Retrieved from Cognitive Rigor, Complexity, Deeper Teaching and Learning, Depth of Knowledge (D.O.K.): https://maverikeducation.com/blog/f/depth-of-knowledge-or-extent-of-learning

Mayring, P. (2000, June). Qualitative Content Analysis. *FQS, 1*(2).

Mislevy, R. J., & Haertel, G. D. (2007, April 7). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice, 26*, 6-20. doi: https://doi.org/10.1111/j.1745-3992.2006.00075.x

Mumford, M. D., Marks, M. A., Connelly, M. S., Zaccaro, S. J., & Johnson, J. F. (1998). Domain-Based Scoring of Divergent-Thinking Tests: Validation Evidence in an Occupational Sample. *Creativity Research Journal, 11*(2), 151-168.

Muss, R. E. (1967, October). JEAN PIAGET'S COGNITIVE THEORY OF ADOLESCENT DEVELOPMENT . *Social Science Premium Collection , 2*(7), 285.

National Education Association. (2020, June 23). *ESSA and Testing: Less testing=more learning*. Retrieved from National Education Association: https://www.nea.org/resource-library/essa-and-testing

New Meridian Corp. (2018). *Ensuring Every Student Succeeds*. Retrieved from New Meridian: https://resources.newmeridiancorp.org/

OSPI (Ed.). (2017, October 30). *Elementary and Secondary Education Act*. (OSPI, Producer) Retrieved 23 2018, Aoril, from Office of the Superintendent of Public Instruction State of Washington: http://www.k12.wa.us/esea/

PARCC. (n.d.). *Partnership For Assessment of Readiness for College and Careers*. Retrieved from https://parcc-assessment.org

PARCC. (n.d.). *PARTNERSHIP FOR ASSESSMENT OF READINESS FOR COLLEGE AND CAREERS* . Retrieved from https://parcc-assessment.org

PARCC. (n.d.). *Proposed Sources of Cognitive Complexity in PARCC Items and Tasks: Mathematics (Summary)* . PARCC.

PARCC, ETS, PEARSON, Assessment SIG Business Meeting . (2014, April). Development of Cognitive Complexity Measures for PARCC . *annual meeting of the American Educational Research Association* . Philadelphia, PA.

PARTNERSHIP FOR ASSESSMENT OF READINESS FOR COLLEGE AND CAREERS. (n.d.). *About*. Retrieved April 18, 2018, from PARTNERSHIP FOR ASSESSMENT OF READINESS FOR COLLEGE AND CAREERS: https://parcc-assessment.org/about/

Pellegrino, J. W., Chudowsky, E., Naomi, E., & Glaser, R. (2001). Knowing What Students Know: The Science and Design of Educational Assessment. *National Academy Press*, 379. Retrieved from http://www.nap.edu.

Pelligrino, J. W., & Wilson, M. (2015). Assessment of Complex Cognition: Commentary on the Design and Validation of Assessments. *The College of Education and Human Ecology, The Ohio State University*, 1543-0421.

Puran, R., Behzadi, M. H., Shavarani, A., & Lotfi, F. H. (2017). The Effects of Training and Other Factors on Problem Solving in Students . *European Journal of Contemporary Education*, 448-460 .

Richardson, V. (1998, September). *National Center for the study of Adult Learning and Literacy*. Retrieved from How Teachers Change: What will lead to change the most benefits student learning?: http://www.ncsall.net/index.html@id=395.html

Roach, A. T., Elliot, S. L., & Webb, N. L. (218-231). Alignment of an alternate assessment with state academic standards: Evidence for the content validity of the Wisconsin Alternate. *The Journal of Special Education, 38*(4). doi:DOI:10.1177/00224669050380040301

Simon, M., & Goes, J. (n.d.). *Dissertation Recipes*. Retrieved from RELIABILITY AND VALIDITY IN QUALITATIVE STUDIES: https://www.dissertationrecipes.com/reliability-validity-qualitative-studies/#:~:text=Validity%20in%20qualitative%20research%20indicates,research%20(Golafshani%2C%202003).&text=Establishing%20validity%20can%20present%20challenges,Cho%20%26%20Trent%2C%2020

Solis-Stovall, L. A. (2020). An Analysis of the Higher Order Thinking Requirements of PARCC Practice Assessments in Grades 3 and 4. *Seton Hall University*.

State of New Jersey Department of Education. (2016, July). *Historical Context: Overview of New Jersey's Statewide Testing Program*. Retrieved from State of New Jersey Department of Education: https://www.nj.gov/education/assessment/history.shtml

Strauss, V. (2016, August 3). *Can Common Core reading tests ever be fair?* . Retrieved March 8, 2018, from Washington Post – Blogs; Washington : https://search.proquest.com/docview/1808556766?accountid=13793

Strobel, J., & Berneveld, A. v. (2009, March 24). When is PBL More Effective? A Meta-synthesis of Meta-analyses. *Interdisciplinary Journal of Problem-Based Learning, 3*(1), 55. doi:https://doi.org/10.7771/1541-5015.1046

Su, H. F., Ricci, F. A., & and Mnatsakanian, M. (2016). When Two Wrongs Made A Right: A Classroom. *Transofrmations, 1*(1). Retrieved from https://nsuworks.nova.edu/transformations

Takahashi, T. (2006, August 1). *Multidimensional Learning*. Retrieved April 20, 2018, from Just Enough and Nothing More: https://justenough.wordpress.com/2006/08/01/multidimensional-learning/

Tarek M. Abdelhamid, M. (1999, December 13). The Multidimensional Learning Model: A Novel Cognitive Psychology-Based Model for Computer Assisted Instruction in Order to Improve Learning in Medical Students. *Medical Education Online, 4*(1).

Tienken, C. (2015, January 25). *PARKING the RHETORIC on PARCC*. Retrieved from https://christienken.com/: https://christienken.com/2015/01/24/parking-the-rhetoric-on-parcc/

Tienken, C. (2017). *Defying Standardization.* London: The Rowman & Littlefield Publishing Co.

Tienken, C. H. (2012, 10 31). The Common Cote State Standards: The Emperor Is Still Looking for His Clothes. *Kappa Delta Pi Record*, 48:4, 152-155. doi:10.1080/00228958.2012.733928

Wasserberg, M. J., & Rottman, A. (2016). Urban High School students' Perspective on Test-Centered Curriculum. *AMerican Secondary Education*, 44(3), 56-71. Retrieved from http://ezproxy.library.berkeley.org/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Furban-high-school-students-perspectives-on-test%2Fdocview%2F1811919870%2Fse-2%3Faccountid%3D38129

Webb, N. L., Alt, M., Ely, R., Cormier, M., & Vesperman, B. (2005, December). The Webb Alignment Tool: Development, Refinement and Dissemination. *Six Reports*.

Webb's Depth of Knowledge Guide. (2009). Retrieved from https://www.aps.edu/sapr/documents/resources/Webbs_DOK_Guide.pdf

Welner, K. G., & Mathis, W. J. (2015, February 12). *Reauthorization of the Elementrary and Secondary Education Act: Time to Move Beyond Test-Focused Policies*. Retrieved from National Education Policy Center: http://nepc.colorado.edu/publication/esea

Wilson, L. O. (n.d.). *The Second Principle*. Retrieved from Understanding the Revised Version of Bloom’s Taxonomy : https://thesecondprinciple.com/essential-teaching-skills/blooms-taxonomy-revised/

Wolf, M. K., Wang, Y., & Huang, B. H. (2014). INVESTIGATING THE LANGUAGE DEMANDS IN THE COMMON CORE STATE STANDARDS FOR ENGLISH LANGUAGE LEARNERS A Comparison Study of Standards. *Middle Grades Research Journal , 9*(1), 35-52.

Wren, D., & Cashwell, A. (2018, February). Measuring Critical Thinking and Problem Solving. *Educational Leadership*.

www.MDE.K12.MS.us. (2009). Webb's Depth of. *Career and Technical Education Definitions*. Retrieved from H T T P : / / WWW . MDE . K 12. M S . U S ;H T T P : / / R E D E S I G N . R C U . M S S T A T E . EDU

Wyse, A. E., & Viger, S. G. (2011, December). How Item Writers Understand Depth of Knowledge. *Educational Assessment, 16*(4), 185-206 .

Wyse, A. E., & Viger, S. G. (2011). How Item Writers Understand Depth of Knowledge. *Educational Assessment, 16*(4), 185-206. doi:10.1080/10627197.2011.634286

Yuan, K., & Vi-Nhuan, L. (2014). *Measuring Deeper Learning Through Cognitively Demanding Test Items ; Results from the Analysis of Six National and International Exams.* William and Flora Hewlett Foundation. Rand.

Appendix A: Coding Categories

[1,1]: Webb's Level 1, Bloom's Level 1. Recall, recognize, or locate basic facts, ideas, principles. Recall or identify conversions between representations, numbers, or units of measure. Identify facts/details in texts.

[1,2]: Webb's Level 1, Bloom's Level 2. Compose and decompose numbers. Evaluate an expression. Locate points (grid, number line). Represent math relationships in words, pictures, or symbols. Write simple sentences. Select appropriate word for intended meaning. Describe/explain how or why. *The two coders agreed that performance tasks asking students to 'describe how or why' must use literal comprehension and verbatim responses.*

[1,3]: Webb's Level 1, Bloom's Level 3. Follow simple/routine procedure (recipe-type directions). Solve a one-step problem. Calculate, measure, apply a rule. Apply an algorithm or formula (area, perimeter, etc.). Represent in words or diagrams a concept or relationship. Apply rules or use resources to edit spelling, grammar, punctuation, conventions.

[1,4]: Webb's Level 1, Bloom's Level 4. Retrieve information from a table or graph to answer a question. Identify or locate specific information contained in maps, charts, tables, graphs, or diagrams.

[1,6]: Webb's Level 1, Bloom's Level 6. Brainstorm ideas, concepts, or perspectives related to a topic or concept. *The two coders agreed that brainstorming in this category requires the recall of prior knowledge and does not include original thought.*

[2,2]: Webb's Level 2, Bloom's Level 2. Specify and explain relationships. Give nonexamples/ examples. Make and record observations. Take notes, organize ideas/data. Make basic inferences or logical predictions from data or texts. Identify main ideas or accurate generalizations. *The two coders agreed that additional thought is not required in developing predictions but relies on prior knowledge. In addition, inferences and predictions in this category have one clear correct answer.*

[2,3]: Webb's Level 2, Bloom's Level 3. Select a procedure according to task needed and perform it. Solve routine problem applying multiple concepts or decision points. Retrieve information from a table, graph, or figure and use it to solve a problem requiring multiple steps. Use models to represent concepts. Write paragraph using appropriate organization, text structure, and signal words. *The two coders agreed that paragraphs written in this category are done in a procedural sense based on the writing process.*

[2,4]: Webb's Level 2, Bloom's Level 4. Categorize, classify materials.
Compare/contrast figures or data. Select appropriate display data. Organize or interpret (simple) data. Extend a pattern. Identify use of literary devices. Identify text structure of paragraph. Distinguish relevant/irrelevant information, fact/opinion.

[2,6]: Webb's Level 2, Bloom's Level 6. Generate conjectures or hypotheses based on observations or prior knowledge. *The two coders agreed that this category is not yet considered higher level thinking, so there is an emphasis on prior knowledge for the performance tasks.*

[3,2]: Webb's Level 3, Bloom's Level 2. Explain, generalize, or connect ideas using supporting evidence. Explain thinking when more than one response is possible. Explain phenomena in terms of concepts. Write full composition to meet specific purpose. Identify themes. *The two coders agreed that this category is considered higher level, so performance tasks in this category do not have obvious answers and, instead, require students to pull from other sources and develop original ideas.*

[3,3]: Webb's Level 3, Bloom's Level 3. Use concepts to solve non-routine problems. Design investigation for a specific purpose or research question. Conduct a designed investigation. Apply concepts to solve non-routine problems. Use reasoning, planning, and evidence. Revise final draft for meaning or progression of ideas.

[3,4]: Webb's Level 3, Bloom's Level 4. Compare information within or across data sets or texts. Analyze and draw conclusions from more complex data. Generalize a pattern. Organize/interpret data, complex graph. Analyze author's craft, viewpoint, or potential bias. *The two coders agreed that performance tasks in this category require students to draw conclusions from more complex data and/or from multiple sources. In addition, in the performance task of 'analyzing author's craft' students must understand how it affects the interpretation of the reading selection.*

[3,5]: Webb's Level 3, Bloom's Level 5. Cite evidence and develop a logical argument for concepts. Describe, compare, and contrast solution methods. Verify reasonableness of results. Justify conclusions made. *The two coders agreed that the emphasis on justification and explaining is an important component of this category.*

[3,6]: Webb's Level 3, Bloom's Level 6. Synthesize information within one source or text. Formulate an original problem, given a situation. Develop a complex model for a given situation. *The two coders agreed that students must develop an original idea through the practice of synthesis.*

[4,2]: Webb's Level 4, Bloom's Level 2. Explain how concepts or ideas specifically relate to other content domains or concepts. Develop generalizations of the results obtained or strategies used and apply them to new problem situations.

[4,3]: Webb's Level 4, Bloom's Level 3. Select or devise an approach among many alternatives to solve a novel problem. Conduct a project that specifies a problem, identifies solution paths, solves the problem, and reports results. Illustrate how multiple themes (historical, geographic, social) may be interrelated. *The two coders agreed that this level may appear in the questions asked in the online-based program, but this may be the highest level that the program can provide.*

[4,4]: Webb's Level 4, Bloom's Level 4. Analyze multiple sources of evidence or multiple works by the same author, or across genres, or time periods. Analyze complex/abstract themes. Gather, analyze, and organize information. Analyze discourse styles.

[4,5]: Webb's Level 4, Bloom's Level 5. Gather, analyze, and evaluate relevancy and accuracy. Draw and justify conclusions. Apply understanding in a novel way, provide argument or justification for the application.

[4,6]: Webb's Level 4, Bloom's Level 6. Synthesize information across multiple sources or texts. Design a model to inform and solve a real-world, complex, or abstract situation.

Taken from :
Hess, K. K., Jones, B. S., Carlock, B. S., & Walkup, J. R. (2009, March). Cognitive Rigor: Blending the Strengths of Bloom's Taxonomy and Webb's Depth of Knowledge to Enhance Classroom-level Processes.

Solis-Stovall, L. A. (2020). An Analysis of the Higher Order Thinking Requirements of PARCC Practice Assessments in Grades 3 and 4. *Seton Hall University*.

Appendix B: Hess Cognitive Rigor Matrix (Reading CRM)

Hess, K. (2013). Linking research with practice: A local assessment toolkit to guide school leaders. Retrieved from https://www.karin-hess.com/cognitive-rigor-and-dok

Appendix C: Hess Cognitive Rigor Matrix (Math-Science CRM)

Hess, K. (2013). Linking research with practice: A local assessment toolkit to guide school leaders. Retrieved from https://www.karin-hess.com/cognitive-rigor-and-dok

Appendix D: Hess Cognitive Rigor Matrix (Writing/ Speaking CRM)

Hess, K. (2013). Linking research with practice: A local assessment toolkit to guide school leaders. Retrieved from https://www.karin-hess.com/cognitive-rigor-and-dok

Appendix E: Coding Table PARCC Geometry

	A,1	A,2	A,3	A,4	B,2	B,3	B,4	B,6	C,2	C,3	C,4	C,5	C,6	D,2	D,3	D,4	D, 5	D,6
1			1															
2			1															
3			1															
4					1													
5					1													
6	1																	
7	1																	
8					1													
9					1													
10	1																	
11					1													
12					1													
13									1									
14					1													
15					1													
16			1															
17			1															
18					1													
19					1													
20						1												
21			1															
22			1															
23		1																
24					1													
25					1													
26					1													
27			1															
28					1													
29		1																
30					1													
31			1															
32			1															
33					1													
34					1													
35						1												
36						1												
37					1													
38						1												
39			1															

Appendix F: Coding Table 10th Grade PARCC English Language Arts

	A,1	A,2	A,3	A,4	B,2	B,3	B,4	B,6	C,2	C,3	C,4	C,5	C,6	D,2	D,3	D,4	D, 5	D,6
1						1												
2		1																
3					1													
4						1												
5	1																	
6		1																
7						1												
8		1																
9					1													
10						1												
11						1												
12									1									
13		1																
14		1																
15	1																	
16		1																
17					1													
18		1																
19					1													
20		1																
21		1																
22		1																
23						1												
24		1																
25	1																	
26		1																
27					1													
28		1																
29						1												
30					1													
31		1																
32		1																
33		1																
34		1																
35					1													
36							1											
37											1							
38		1																
39							1											
40					1													
41					1													
42					1													
43					1													
44					1													
45					1													
46										1								
47		1																
48		1																
49		1																
50		1																
51						1												
52					1													
53					1													
54					1													
55					1													
56					1													
57									1									
58					1													