An Analysis of the Higher Order Thinking Requirements of PARCC Practice Assessments in Grades 3 and 4

Leslie A. Solis-Stovall
leslie.solis@student.shu.edu

An Analysis of the Higher Order Thinking Requirements of PARCC Practice Assessments in Grades 3 and 4

Leslie A. Solis-Stovall

Dissertation Committee

Christopher Tienken, Ed.D
Lavetta Ross Ed.D
Samuel Fancera 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
2020

COLLEGE OF EDUCATION AND HUMAN SERVICES
SETON HALL UNIVERSITY

APPROVAL FOR SUCCESSFUL DEFENSE

Leslie Solis has successfully defended and made the required modifications to the text of the doctoral dissertation for the Ed.D. during this **Summer Semester 2020**.

DISSERTATION COMMITTEE
(please sign and date beside your name)

Mentor:
Christopher Tienken ______ 5-15-20
Date

Committee Member:
Samuel Fancera ______ 5-15-20
Date

Committee Member:
Lavetta Ross ______ 5-15-20
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 to be bound as page number two.

Abstract

This mixed methods study aimed to categorize and analyze the frequencies and percentages of complex thinking in the PARCC practices assessments grades 3 and 4. The Hess' Cognitive Rigor Matrix was employed for the first part of the study to code each of the PARCC assessment questions in Language Arts and Mathematics Grades 3 and 4 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 and Mathematics PARCC Practice Tests in grades 3 and 4:

1. The questions in the Language Arts PARCC tests in Grades 3-4 were rated at an overall higher percentage for lower level questions.
2. The questions in the Mathematics PARCC tests in Grades 3-4 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 the culmination of the journey in gaining my Ed.D, which was just like climbing Mount Everest. Accompanied with encouragement, hardship, trust, and frustration this doctoral journey has truly been an experience of a lifetime. Throughout my educational endeavors, I have been blessed with the love and support of my family. For that I am grateful and want to thank them from the bottom of my heart. I want to give a special thanks to my parents, who have instilled in me the drive needed to succeed in anything I set my mind and heart to. My mother for loving me unconditionally and instilling in me the positive qualities that make me the person I am today. Perseverance, kindness, confidence are all things my mom has strived to infuse in my life. I don't know where I would be without you mom. Love you!

I especially dedicate this dissertation to my loving father whom I lost on April 8, 2020 due to the ongoing COVID-19 pandemic. I am truly in disbelief. Just a couple of weeks ago he kept asking me when I was going to finish. He always had my best interest at heart and reveled in the success of my sister and I. Dad, I thank you for always guiding me and supporting me in anyway you possibly could. God knows how much I love you and miss you. I am certain I have made you proud and that brings me peace.

Thanks to my sister Alice and my nephew Mason for their selfless love, care and dedicated efforts which contributed a lot for the completion of this dissertation. Love you both immensely.

I owe thanks to a very special person in my life, my husband, Omar for his continued and unfailing love, support and understanding during my pursuit of my Ed.D, and for making the completion of this degree possible. You were always around at times I thought that it was impossible to continue, you helped me to keep things in perspective while also taking care of the kids in my long days of writing and the long weekends spent at Seton Hall. Thank you for believing in me my love.

To my three heartbeats, my sons Ayden and Liam, and my daughter Scarlett. Thank you for being patient with mommy during those long days. I know it has been tough. My greatest accomplishment is having you guys. I am so lucky and blessed to have such wonderful, caring, and understanding children. Mommy loves you forever and a day.

At this moment of accomplishment I am greatly indebted to my committee. To my mentor Dr. Tienken, this work would not have been possible without your guidance and involvement, your support and encouragement on a daily basis. Under your tutelage I successfully overcame many difficulties and learnt a lot. Your zeal for perfection and passion for education is truly inspiring and I thank you for your unwavering support. My earnest thanks to Dr. Fancera and Dr. Ross, I am grateful for the valuable advice, constructive criticism, positive appreciation and counsel throughout the course of my research which led to the successful completion of my dissertation. Lastly, I greatly appreciate and acknowledge the support received from Dr. Sydoruk, thank you for taking the time to advise me during the methodological stage of my dissertation.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

The preparation of students for the increased education demands regarding the attainment of 21st century skills necessary for economic viability in a global economy has become a priority in many state education policies in the United States. According to the World Economic Forum (2015), to thrive in today's innovation-driven economy, workers need a different mix of skills than in the past; in addition to foundational skills like literacy and numeracy, they need competencies like collaboration, creativity, problem solving, and dispositions like persistence, curiosity, and initiative.

There is rhetorical emphasis on the part of education policy makers in developing and assessing the complex thinking and critical thinking skills of students in response to what some see as a shift to a knowledge economy. Officials at the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium made claims that the PARCC assessments measured students' readiness to master rigorous academic content at each grade level, think critically and apply knowledge to solve problems; with the main goal of preparing students for college and career readiness (PARCC, 2018). With the demise of the PARCC consortium in almost all states as of 2020, the name PARCC is dead, but the questions and PARCC-like tests live on, as many former PARCC states now contract with the company that bought the PARCC questions, New Meridian Corporation, and purchase test questions from that company. The tests are PARCC tests, with a different name.

Higher Order Thinking

A single definition of higher-order thinking does not exist. Newmann, (1988) suggests a simple definition: higher order thinking signifies challenge and expanded use of the mind; lower order thinking signifies routine, mechanistic application and constraints on the mind (p. 2). Challenge, or the opportunity to expand the use of mind, occurs when a person must interpret, analyze, or manipulate information, because a question to be answered or a problem to be solved cannot be resolved through the routine application of previously learned knowledge (p. 2). Many terms and theories have been used to define higher order thinking: critical thinking; complex thinking; abstract, deductive, and inductive reasoning; formal and informal reasoning; Piaget's formal operational thought; Kohlberg's post-conventional moral stages; metacognition; Bloom's categories of knowledge, comprehension, application, analysis, synthesis, and evaluation; Webb's DOK levels analyzing the cognitive complexities of content; Hess' Cognitive Rigor Matrix superimposing Blooms Taxonomy, and Webb's DOK to describe rigor and deeper learning in the fields of education and assessment (Bloom, 1956; Hess et al., 2009a 2009b; Newmann, 1988). Regardless of the ambiguities surrounding the term, it has been determined in policy rhetoric that the use of higher order thinking in education is of high importance in both instruction and assessment.

Standardized Testing

Typically, policy makers use high stakes standardized testing policies to enforce compliance with mandated curriculum initiatives. Driven by the No Child Left Behind Act (NCLB, 2002), and Every Student Succeeds Act (ESSA), standardized assessments are mandated in subjects and grade levels. The results from those tests provide data that officials in some states use to determine teacher and school leader evaluation ratings and the desire to achieve high ratings tend to influence instructional leadership practices. According to Ives & Obenchain (2016), high stakes standardized testing narrows curriculum to what is most likely tested. In many cases, high stakes standardized tests, focus on lower level skills and procedures.

Narrowing the curriculum based on the topics most likely tested impedes the instructional strategies that foster complex thinking (Tienken, 2017). Specifically, the PARCC consortium claimed that the PARCC test questions assessed higher order thinking, and more specifically, complex thinking. Officials stated that their, "PARCC's Cognitive Complexity Framework guides item development and helps determine the overall complexity of the ELA and math tasks in the assessments, while serving as a bridge between the standards (PARCC, 2018). The frameworks are claimed to serve as a bridge between the standards and these tools in a variety of ways, including by clarifying areas of emphasis in each grade and what changes in the standards from one grade to the next (PARCC, 2018).

Problem Statement

Officials at the Partnership for Assessment of Readiness for College and Careers (PARCC, 2018) claimed that the PARCC assessments were an "educational GPS system," assessing students' current performance, and pointing the way to what students need to learn to be ready for the next grade level and, by high school graduation, for college and/or a career. They also claim that the assessments are designed to give schools and teachers more information to improve and enhance instruction (PARCC, 2018).

Many of PARCC's (2020) assessment design reflect the partnership's ambitions to meet these high expectations for next-generation, college and career readiness assessments. The PARCC were originally adopted by 19 states, however, as of 2019 less than a handful six states use the original PARCC assessment. However, many states that stopped using the original

PARCC test use tests that are developed from the PARCC item banks. Several states, like Massachusetts that originally used PARCC moved away from the name of PARCC, but the new Massachusetts Comprehensive Assessment System (MCAS) retains many of the multi-step PARCC-like questions because the authors of new MCAS, Measured Progress (2016), purchase items from the PARCC item bank managed by Pearson. Officials at the New Jersey Department of Education renamed their assessments but still use the PARCC test items purchased from the New Meridian Corporation (Eno, 2018). The MCAS, and other similar state assessments, like those in New Jersey are PARCC tests with a different name.

Assessing Thinking

Some researchers have deemed the assessment of complex thinking important to ensure preparation of students for global competiveness (e.g., Bechard, Hess, Camacho, Russell, & Thomas, 2012; IBM, 2012; National Academy of Sciences, 2007, 2010; Tienken, 2017; World Economic Forum, 2015, Zhao, 2012). As of 2018, only three states used "pure" PARCC test: Maryland, and New Mexico, plus the District of Columbia (Sawchuk, 2018). Illinois was the most recent state to only adopt some parts of the assessment including "complex writing tasks that require strategic reasoning and extended investigation to solve problems" (Sawchuk, 2018).

The existing literature on the topic of evaluating the complex thinking requirements of the PARCC is limited, and thorough investigations of the assessments complex thinking requirements have not taken place. Some studies used Webb's Depth of Knowledge to examine the complexity of nationally and internationally administered tests (Kun & Vi-Nhuan, 2012), but a dearth of evidence exists on the complex thinking requirements of PARCC at all grade levels (Tienken, 2015).

Some have questioned the validity of PARCC's claims of complexity and its methods for determining cognitive complexity (Gewertz, 2019; Tienken, 2019; Yuan & Le 2014). Yuan & Le (2014) employed Webb's DOK and the PARCC framework to examine the cognitive demand of six nationally and internationally administered tests. Yuan and Le (2014) determined that PARCC framework provided guidelines to create an overall complexity score, but opted to deviate from the recommended scoring mechanism. The scoring rubric gave relatively greater weight to the difficulty of the content and relatively less weight to the cognitive processes, and found that this approach did not work well for open-ended items (Yuan & Le, 2014 p. xii). PARCC's methods for determining cognitive complexity differ from the generally accepted definition of cognitive complexity.

According to Tienken (2015, p. 5), "PARCC tests are simply measuring 19th century skills with a 20th century tool (computer)", because of its alignment to CCSS. The CCSS mandate knowledge and skills that have not significantly changed over the last 150 years. Some commentators from business, economics, and education circles argue that the types of higher-order thinking skills that students need to be globally competitive include creative thinking and strategic thinking (Sforza, Tienken, & Kim, 2016).

There are no current studies at the elementary grade levels that test the claim that PARCC questions require complex thinking, and thus school leaders have no way of knowing the meaningfulness of the results as they relate to student higher order thinking.

Purpose of the Study

My purpose for this Mixed-methods study was to describe the way(s) in which 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 Grades 3 and 4 associate with the language that promotes higher-order thinking found in research literature. The Elementary grades were selected for this study due to the lack of existing research.

Research Questions

The study was grounded by an overarching research question: What types of thinking are assessed by the questions on 2016 PARCC practice tests in English language arts and mathematics in grades 3 and 4?

1. In what way(s) does the language of the questions on the English language arts section of 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 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 mathematics section of 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 associate with the language that promotes higher-order thinking found in research literature?
3. What is the distribution of thinking on the 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in English language arts and mathematics in Grades 3 and 4?

Methodology Overview

I conducted a Mixed-methods study with qualitative and quantitative content analysis methods to describe the way(s) in which the language of the English language arts and mathematics questions on the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 associates with the language found to promote higher-order thinking found in research literature. I used Hess's Cognitive Rigor Matrix (Hess, 2009b) framework to categorize the language for each question of the assessments. The study used two analysts in coding the language of each question on each test (Merriam, 2009, p. 216). A double-rater read behind consensus model was used based on the work of Miles, Huberman, and Saldaña (2014, p. 84) to increase reliability. Furthermore, the analysts used Mayring's Step Model with associated coding protocols to structure the process.

For the current study, all 220 reading, writing and mathematics questions were chosen from the Grades 3 and 4 English language arts and mathematics PARCC practices tests. Some questions in these assessments had two parts and were coded accordingly. Two coders reviewed the assessment questions and compared the language of the questions to the language of higher-order thinking found in the literature, as represented on the Hess' Cognitive Rigor Matrix.

The overarching aim of qualitative research is to gain an understanding of certain social phenomena (Renz, Carrington, & Badger, 2018). This study used content analysis methods to test PARCC's claims of critical thinking, and to describe the complex thinking requirements of the English language arts and mathematics in the PARCC practice tests in Grades 3 and 4. Specifically Weber (1990) explained, how qualitative content analysis extends beyond the mere counting of words while seeking to provide knowledge and understanding of the phenomena under study through systematic method of coding and classifying text to reveal patterns and themes. Quantitative methods were utilized in the second part of the study to describe the differences and similarities of higher order thinking that exist in the language of the assessment. This study employed the sequential exploratory strategy, which involved a first phase of qualitative data collection and analysis, followed by a second phase of quantitative data collection and analysis that builds the results of the qualitative phase (Cresswell, 2007). Consequently, The purpose of this strategy is to use quantitative data and results to assist in the interpretation of qualitative findings (p. 211).

Conceptual Framework

The conceptual framework utilized in this study was Hess' Cognitive Rigor Matrix. The Cognitive Rigor Matrix (CRM) enhances the instructional, assessment planning and practices (Hess, Carlock, Jones, & Walkup, 2009a). The CRM superimposes two cognitive complexity frameworks: Bloom's Taxonomy and Webb's Depth of Knowledge (Hess, 2013) The CRM is a two dimensional framework that incorporates the revised Blooms Taxonomy and Webb's DOK. Questions and prompts can be categorized by the cognitive skills required, such as recall, comprehension, or analysis and the type of thinking facilitated, be it higher level or lower level (Simpson et al., 2015).

The CR matrix consists of 24 cells; namely, 6 levels of the revised Bloom's Taxonomy (horizontally) dichotomized into 4 levels of DOK (vertically) (Hess et al., 2009a). The lowest level of the CR matrix represents the association between Bloom's *remember/recall* and Webb's Level 1: recall and reproduction, CR cell [1,1]. The objective for this cell is to recognize basic details and facts through rote memorization. The connection between Bloom's highest level create and Webb's Level 4 extended thinking, CR cell [4,6] expects students to be able to synthesize information across multiple sources. Moreover, they need to be able to apply learnt information into a design and solve real, abstract, complex situations (Hess et al. 2009b). Hess developed the CRM due to Bloom's Taxonomy limitations regarding its verb indicators when analyzing curriculum and assessments. Webb's depth of knowledge (DOK) provides a framework to analyze the language of test questions in relation to the language of higher order thinking (Hess et al., 2009a)

Significance of the Study

There have been previous studies that have used the Hess's framework to determine the complex thinking requirements in curricula resources (Sydoruk, 2018). This study extends that work to a nationally used standardized test. No comparable studies have sought to use the Hess Cognitive Rigor Matrix to determine the complex thinking requirements of PARCC. Many 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). This allowed educators to effectively analyze and differentiate tasks in the curriculum, hence enabling them to create effective lessons. I sought to extend that work by using Hess' Cognitive Rigor Matrix (CRM) to describe the complex thinking requirements from a nationally recognized standardized test. Determining the complex thinking embedded in the Partnership for Assessment of Readiness for College and Career (PARCC) allows educational stakeholders to gain knowledge and information that can be used to prepare students for the challenges of the 21st century.

Hess Cognitive Rigor Matrix affords the opportunity to systematically examine the extent in which higher learning is embedded in the PARCC elementary tests. This model was used because of it s application of two renowned scientific frameworks Bloom's Revised Taxonomy and Webb's Depth of Knowledge. The Bloom's Revised Taxonomy is comprised of two dimensions, knowledge, and cognitive processes. The taxonomy table employs six categories: Remember, Understand, Apply, Analyze, Evaluate, and Create. The use of the Taxonomy Table analyzes and provides an indication of the extent to which more complex kinds of knowledge and cognitive processes are involved (Krathwohl, 2002). Webb's DOK provides a method of interpretation by assigning DOK levels to objectives within standards and assessments as a form of analysis (Webb, 2002). The DOK levels are: DOK-1 Recall & Reproduction, DOK-2 Basic Application of Skills/Concepts, DOK-3 Strategic Thinking, and DOK-4 Extended Thinking. Both frameworks differ in application; Bloom's Taxonomy categorizes the cognitive skills (Hess et al., 2009b). This in turn describes the type of thinking processes necessary to answer a question. Webb's Depth of Knowledge model relates more closely to the depth of content understanding and scope of a learning activity (Hess et al., 2009b). Hess (2006) superposed Bloom's Taxonomy and Webb's DOK due to lack of one-to-one correspondence developing the Cognitive Rigor Matrix. With the use of CRM, I sought to analyze complex thinking in a nationally used standardized test.

The study extended the research on the application of higher order thinking in assessments; the analysis of this study was aimed at the elementary level, specifically focusing on Grades 3 and 4. It is important to examine these assessments that are meant to measure student achievement of CCSS because, in addition to carrying high stakes, mandatory testing is known to impact instruction in unintended ways, such as narrowing the curriculum to what is most likely tested (Polleck & Jeffrey, 2017). If what is most likely tested is lower level thinking, then the curriculum could be narrowed to that type of thinking.

Limitations

There are several limitations that should be noted regarding this particular study. In this case study the results are only applicable to this case and results should not be generalized to any other contexts or assessments. The results only apply to the practice questions released on (Spring 2016) from parcc-assessment.org. Furthermore, two coders were trained using Hess' Cognitive Rigor Matrix; the results are based on their expertise, experience, and perceptions of what complex thinking skills are. Additionally, the developer of the CRM may have biases that I am unaware of. Lastly, my study was limited to only analyzing the practice test due to the availability of questions.

Delimitations

Delimitations of this study include the grade levels selected. For this study, I chose to focus on Grades 3 and 4 of this assessment due to my experience as an elementary educator. Results cannot be generalized to other grade levels. Moreover, the practice questions are not current and were developed between the years 1998 -2015 noted on the website. Furthermore, the numbers of coders are considered delimitation due to bias. In addition, the two coders and I agreed that due to the limitations in the availability of the PARCC assessments, the practice tests would be used. The PARCC practices test include past questions that can be coded to determine the complexity of the current assessment.

Definitions of Terms

Cognitive complexity is a multidimensional phenomenon that "indexes the degree of differentiation, articulation, and integration within a cognitive system; individuals with more "developed" cognitive systems have more differentiated (i.e., numerically larger), more abstract (i.e., more refined or specialized elements), and more integrated (i.e., more organized) construct systems (Da'as, Schechter, & Qadach, 2018).

Common Core State Standards (CCSS) is a set of academic standards in mathematics and English language arts. These learning goals outline what a student should know and be able to do at the end of each grade (CCSS Initiative, 2018).

Higher order thinking skills distinguishes two contexts in which these skills are employed: contexts where the thought processes are needed to solve problems and make decisions in everyday life; and contexts where mental processes are needed to benefit from instruction, including comparing, evaluating, justifying and making inferences (Wheeler & Haertel, 1993 as cited in Forster, 2004).

Partnership for Assessment of Readiness for College and Career (PARCC) is a group of states working together to develop a set of assessments that measure whether students are on track to be successful in college and careers (PARCC, 2019).

Practice test is a form of assessment used for familiarizations of the kinds of items and format used in the actual assessments (PARCC, 2019).

Organization of the Study

In Chapter I, I provided an overview of the problem related to PARCC's claims of cognitive complexity in accordance to HESS's Cognitive Rigor Matrix. Existing literature on the topic of evaluating the complex thinking requirements of the PARCC is limited, and investigations of the assessments cognitive complexity have not taken place. It is necessary for studies to validate and attest these claims.

Chapter II is comprised of a literature review on previous research regarding higher order thinking, studies conducted on the cognitive complexity of assessments and curriculum, and an in depth look into PARCC's claims and criticisms.

Chapter III expanded on the methodological approach and procedures for the qualitative study. Data collected from PARCC practice assessments are tested for complexity using Hess' Cognitive Rigor Matrix.

Chapter IV organized and presented the data and main findings of the study.

Chapter V summarized the statistical findings, provided an analysis of data, recommendations for future research, and a conclusion.

Chapter II: Review of Literature

The purpose for this Mixed-methods study was to describe the way(s) in which the 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 3 and 4 associate with the language that promotes higher-order thinking found in research literature. Elementary grades were selected for this study because of the lack of research. The research questions guide my review of the literature and encompass the following sections: higher order and cognitive complexity, PARCC assessments, related studies, and theoretical framework. The review of the literature served as a means of identifying empirical studies on theories of higher order thinking, the measurements of the complex thinking requirements, theoretical frameworks related to the higher order thinking, and a critique of existing studies on topics similar to the this study.

Literature Search Procedures

Utilizing Boote and Biele's (2005) guidelines for literature reviews, peer-reviewed literature was acceessed from multiple online databases including SAGE, ProQuest Databases, EBSCO, and Google Scholar. Key words such as, higher order thinking, cognitive complexity, critical thinking, and PARCC's claims were some of the variables searched and included in my research. 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.

Criteria for Inclusion of Literature

- Peer- reviewed original research
- Dissertations
- Government reports
- Non-peer reviewed surveys of skills desired by multinational corporations,
- Peer and non-peer reviewed literature regarding PARCC Claims
- Theoretical literature
- Seminal works
- Studies published within the last 50 years.

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 were numerous terms in relation to higher order thinking, which made it difficult to yield results. In relation to the terms complex thinking, higher order thinking, cognitive complexity, rigor, critical thinking, strategic thinking, all were often interlaced in studies with no clarification in its definition.

Very little literature exists regarding the validity of PARCC's claims to assess higher order thinking, and a lack of experimental research regarding the complexity of the PARCC assessments. Much of the existing literature examines curriculum and the Common Core Standards in terms of cognitive complexity. Literature regarding curriculum predominantly focuses in the implementation of tasks in the classroom that are developed with levels of hierarchy in the higher-order thinking spectrum. Studies that contained large samples examined statewide initiatives for Common Core State Standards, such as analysis of standards to determine ranges of cognitive complexity required to demonstrate knowledge.

Much of the literature surrounding PARCC assessments is produced by non-peer reviewed outlets, PARCC itself, and think-tanks known to be biased toward standardized testing and supporters of PARCC (Tienken, 2019). Overall, much of the existing literature is focused on promoting the assessment. Studies on the PARCC assessments ability to determine whether they are college and career or on track has not been substantially tested or researched. Similarly, in studies of the PARCC assessment's complexity, the consortium used their own framework to determine the complexity of the questions. This could indicate bias in the methodologies and results.

Review of Literature Topics

The purpose of this case study with Mixed-methods was to compare, analyze, and describe the language of complex thinking embedded within the 2016 PARCC practice assessments in mathematics and language arts Grades 3 and 4. The purpose of the literature review was to critique the existing literature regarding the thinking requirements. The literature review also presented a review of definitions of higher-order thinking. Additionally, this literature review identified frameworks that are in alignment with the coding of the PARCC assessments.

Higher Order Thinking, Complex Thinking and Cognitive Complexity

The mastery of higher order thinking skills is one of the important skills stressed in the educational push for 21st century skills attainment. The rigorous demands include trends such as, "rigorous standards, teacher evaluation and accountability systems based on students' achievement on standardized assessments" (Peterson, 2017, p. 1). In the quest of gaining information regarding higher order thinking, a dearth of information on its application has been found. Although the information can be acquired many have failed to provide a common definition for educational settings.

Newmann (1988), defined higher order thinking as “the interpretation, analysis, or manipulation of information to answer a question that cannot be resolved through the routine application of previously learned knowledge” (p.60). Higher order thinking is a considerable definition but vague, and hard to apply in an educational setting. Newmann’s (1990), later work titled *Higher Order Thinking in Teaching Social Studies: A Rationale for Assessment of Classroom Thoughtfulness,* mentioned research that suggests certain obstacles regarding the application of higher order thinking, problem solving, reasoning, critical thinking, and creative thinking. These obstacles include defining higher order thinking, evaluating student performance in thinking; class size and teaching schedules that prevent teachers from responding in detail to students’ work; curriculum guidelines and testing programs that require coverage of vast amounts of material; students’ apparent preferences for highly structured work with clear answers; and teaches’ conceptions of knowledge that emphasize the acquisition of information more than interpretation, analysis and evaluation (Newmann, 1990). There is an emphasis placed in the many factors that may affect the application of higher order thinking skills in the educational setting. Newmann’s *The Relationship of Classroom Thoughtfulness to Students' Higher Order Thinking: Preliminary Results in High School Social Studies* (1990) discusses the approaches to building curriculum and instruction in a four-step process:

1. Identify the main problems or challenge; that students should be competent to address
2. For each problem, identify the specific body of in-depth knowledge, the cluster of analytic skills, and the main dispositions needed for success in addressing the problem.

3. Experiment with alternative methods for teaching the specific knowledge, skills and dispositions relevant to each problem.
4. Codify the results to produce guidelines for curriculum and pedagogy most likely to assist students in resolving each of the major cognitive challenges identified (p.7)

Although systematic this approach provides an individualize process, giving the educator the ability to assess and differentiate application according to need.

Brookhart's (2010) *How to Assess Higher-Order Thinking Skills in Your Classroom* described higher order thinking as students "being able to relate their learning to other elements beyond those they were taught to associate with it, such as relating the content to prior knowledge or making connections outside of the curriculum (p. 5)." This example indicates that there is much subjectivity into the degree to which a student can master this skill. There is interest and emphasis on higher order thinking, mainly how well students' can apply their thinking ability in various forms. According to Brookhart (2010), higher order thinking fall into three categories: (1) higher order thinking in terms of transfer, (2) those that define it in terms of critical thinking, and (3) those that define it in terms of problem solving.

Higher order thinking in terms of transfer requires students not only to remember but also to make sense of and be able to use what they have learned (Anderson & Krathwohl as citied by Brookhart, 2010). Research has shown that students have struggled with the ability to take learnt knowledge and transfer and apply in a different setting. Two of the most important educational goals are to promote retention and to promote transfer (when it occurs, indicates meaningful learning) (Anderson & Krathwohl, 2001). This approach has informed construction of the cognitive dimension of the revised Bloom's taxonomy (Brookhart, 2010, p.5). The educator's goal is to provide learning opportunities that can be applied in the classroom or in real world situations. In Halpern's article, *Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring* (1998), she mentioned the application of thinking within the knowledge domain, the usual method for teaching content. Educators want students to use in multiple domains because instruction in most courses focuses on content knowledge instead of the transferability of thinking skills. Similarly, on the assessment side, the analysis of cognitive processes is intended to help educators (including test designers) broaden their assessments of learning (Anderson & Krathwohl, 2001). Although assessment tasks that tap recalling and recognizing have a place in assessment, these tasks can (and often should) be supplemented with those that tap the ful1 range of cognitive processes required for transfer of learning. (Anderson & Krathwohl, 2001).

In the book *Critical Thinking and Higher Order Thinking : A Current Perspective* (2012), Shaughnessy discussed the history of critical thinking and how it was traced to the Socratic dialogues in the fifth century; Socrates attempted to help people examine their thoughts and actions through systematic probing and questioning. Critical thinking a term widely used to replace the term higher order thinking is continuously assessed and implemented in classrooms. Critical thinking can be defined as "artful thinking", which includes reasoning, questioning and investigating, observing and describing, comparing and connecting, finding complexity, and exploring viewpoints (Brookhart, 2010 in Barahal, 2008). In comparison to transferring higher order thinking skills in different settings, critical thinkers are predisposed to use these types of skills without prompting; they are expected to apply in their daily lives (Halpern, 1998; Brookhart, 2010; Shaughnessy, 2012). Though most people would agree that critical thinking is important, few can clearly articulate what it is, further it is assumed that it is fostered in our classroom with little evidence to support this view (Shaughnessy, 2012). Critical-thinking skills are often referred to as higher order cognitive skills to differentiate them from simpler (i.e., lower order) thinking skills (Halpern, 1998). Being able to apply this skill in and out of the classroom means "students can apply wise judgment or produce a reasoned critique" (Brookhart, 2010, p. 5). The term critical thinking contains many components that are applied in the cognitive processes and to apply, one must be able to evaluate and reflect.

A problem is a goal that cannot be met with a memorized solution (Brookhart, 2010). Problem solving is considered, as the non-automatic strategizing required for reaching a goal (Nitko & Brookhart, 2007 as cited by Brookhart, 2010). Problem solving is one of the many facets of the term higher order thinking and it tends to overlap with the term critical thinking. Both terms include critical evaluation and choosing the proper approach to solving a problem, a skill that requires reasoning (Brookhart, 2010 as cited by Sydoruk, 2018). Higher-order thinking skills involve critical thinking and learning to solve a problem, which in turn allows you to be challenged and think critically. Today's schools are focused in developing students that are prepared to create new ideas while making complex decisions.

To meet the students' educational needs educators are expected to provide more than the basic skills. Moreover, the complexity of the cognition underlying higher order thinking, shown by the psychological research on these skills, describes the challenges of teaching them (Richland & Begolli, 2016). The various terms "higher order thinking, critical thinking, problem solving, rational thought, and reasoning tend to be confusing" (Lewis & Smith, 1993). Defining thinking skills, reasoning, critical thought, and problem solving have proved to be troublesome and has been referred as a conceptual swamp by practitioners (Cuban, 1984, as cited by Lewis & Smith, 1993).

Despite the challenges of defining higher order thinking, educational stakeholders have agreed with the importance of its application (Bloom, 1956; Dewey, 1933; Hess et al., 2009a 2009b; Lewis & Smith, 1993; Newmann, 1988). Much of the existing literatures embed examples of higher order thinking skills in their studies. According to King, Goodson, and Rohani (1998), higher order thinking skills include critical, logical, reflective, metacognitive, and creative thinking that is activated when individuals encounter unfamiliar problems, uncertainties, questions, or dilemmas. Successful applications of the skills result in explanations, decisions, performances, and products that are valid within the context of available knowledge and experience that promote continued growth in these and other intellectual skills (King, Goodson, & Rohani, 1998). The current push for 21st century skills stresses an educational shift from a need to help students acquire knowledge with much information easily accessible via technological resources, to a focus on the ability to create, innovate, critique, evaluate, and integrate information now available to emerging adults (Richland & Begolli, 2016).

Dewey (1933) and Bloom (1956) both promoted the development of students' critical and analytical abilities, but there is less agreement in its application. Many have emphasized on what can be referred to as critical-analytic thinking, especially the capacity to evaluate multiple streams of information in different representational formats in fundamental content areas, such as English language arts, mathematics, and science (Brown, Afflerbach, & Croninger (2014); National Assessment Governing Board (2010); Common Core State Standards (2010); Next Generation Science Standards (2013). With the intentions of students learning to use cogent reasoning and evidence collection skills that are essential for success in college, career, and life (Common Core State Standards Initiative, 2018). Students' who are successful vis-à-vis these competencies will be metacognitive throughout the performance assessment (Brown, Afflerbach,

& Croninger, 2014). Commentators from the educational, political, and business argue that higher order thinking skills must be embedded in the assessing of students to be globally competitive.

Assessment of Cognitive Domain Frameworks

Bloom's Taxonomy

In 1956, Benjamin Bloom and a group of educators developed a classification of levels of intellectual behavior important in learning (Hess, Carlock, Jones, & Walkup, 2009b). The seminal Taxonomy of Educational Objectives: The Classification of Educational Goals—Handbook I, Cognitive Domain (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) represented years of collaboration by the Committee of College and University Examiners, and was the first of three volumes that together would become known as *Bloom's Taxonomy of Learning* (so named after Benjamin Bloom, the original committee chair) (Irvine, 2017).

The original Bloom's Taxonomy is a six-level classification system containing subcategories, all lying along a continuum from simple to complex and concrete to abstract (Athanassiou, McNett, & Harvey, 2003; Armstrong, n.d.). The taxonomy includes:

- 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 (Athanassiou, McNett, & Harvey, 2003; Armstrong, n.d.):."

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

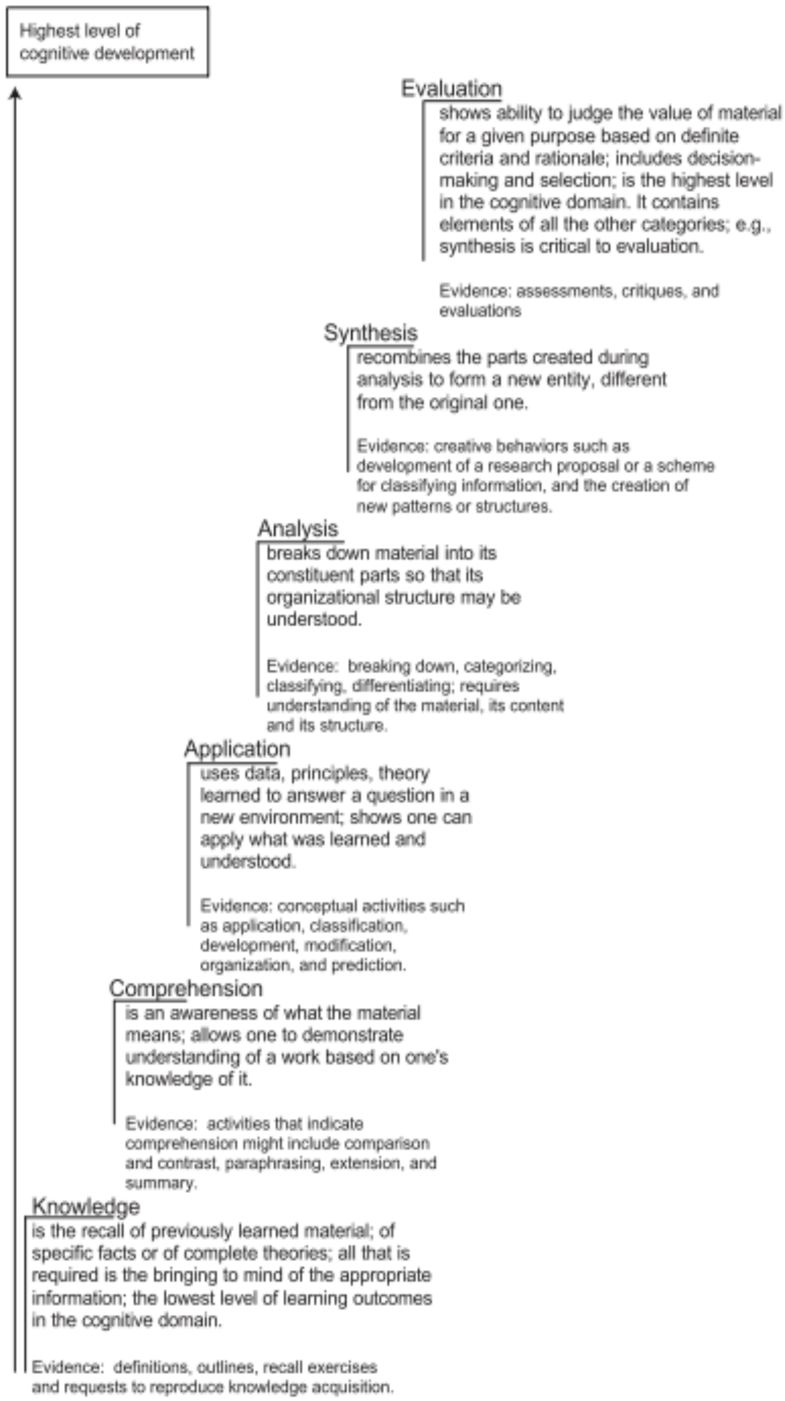

Figure 1. Bloom's Taxonomy of Educational Objectives: Cognitive Domain (Anthanassiou, McNett, & Harvey, 2003).

Blooms taxonomy has been used for the dimensions of effective teaching curriculum analysis (Anthanassiou, McNett, & Harvey, 2003). Moreover, Bloom's Taxonomy helps form educational lessons that develop thinking skills over a range of complexity (Hess, et al., 2009).

Revised Bloom's Taxonomy

Building on the prior taxonomy model, the revised version provides a way to better understand a broad array of assessment models and application (Airasian & Miranda, 2002). Anderson and Krathwohl's revision of Bloom's Taxonomy (2001) is a two-dimensional framework: Knowledge and Cognitive Processes and resembles the subcategories of the original Knowledge category. (See figure 2.)

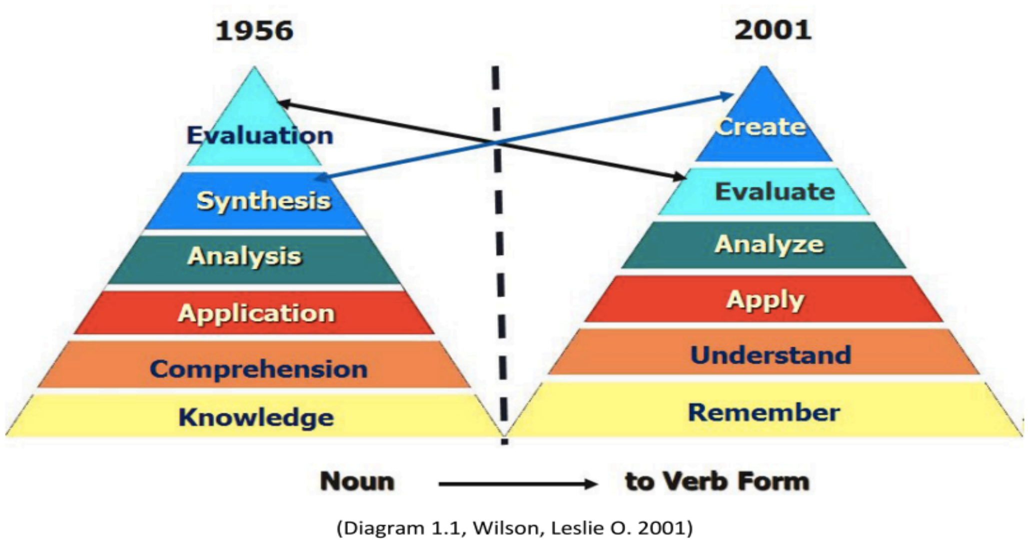

Figure 2. Revised Bloom's Taxonomy (Wilson, 2001).

According to Hess, et al., (2009), the cognitive processes resemble those found in the original taxonomy, but the placement of each level on the taxonomy continuum shifted (e.g., evaluation no longer resides at the highest level) and includes expanded and clarified descriptions for analyzing educational objectives. The latter resembles the six categories of the original Taxonomy with the Knowledge category named Remember, the Comprehension category named

Understand, Synthesis renamed Create and made the top category, and the remaining categories changed to their verb forms: Apply, Analyze, and Evaluate. (Krathwohl, 2002)

The new levels identify cognitive learning (arranged from lower- order to higher-order levels of learning) (IACBE, 2015):

- Remembering – Retrieving, recognizing, and recalling relevant knowledge from long-term memory
- Understanding – Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining
- Applying – Using information in new ways; carrying out or using a procedure or process through executing or implementing
- Analyzing – Breaking material into constituent parts; determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing
- Evaluating – Making judgments based on criteria and standards through checking and critiquing; defending concepts and ideas
- Creating – Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing (IACBE, 2015, pp. 9-10)

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

Figure 3. Comparison of descriptors (Hess et al., 2009).

A study examined six teachers, working at a variety of grade levels, to describe actual instructional units they had taught in their main subject area (Airasian, & Miranda, 2002). According to Airasian, & Miranda, (2002), this provided useful information about the validity of classroom and statewide assessments as evidenced by the alignment of the assessments. According to Hess, "the restructuring of the original taxonomy recognizes the importance of the interaction between content (characterized by factual, conceptual, procedural, and metacognitive knowledge) and thought processes" (Hess, et al., 2009, p2). The focus placed on the Taxonomy Table allowed for an increase in the alignment of assessment with both objectives and instruction

(Airasian, & Miranda, 2002). In addition, the Taxonomy Table can be used to increase the alignment of school-wide or district-wide curriculum and instruction with state standards and state-mandated assessments, which will enable teachers to focus on the standards without "teaching to the test" (Airasian, & Miranda, 2002).

Webb's Depth of Knowledge.

Webb's work has forced states to volte-face the meaning of test alignment to include the intended cognitive demands to which students are expected to demonstrate. 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 et al., 2009b). Webb describes his depth-of- knowledge levels as "nominative" rather than as a taxonomy, meaning that DOK levels describe four different ways a student might interact with content (p2). The identification of Webb's DOK levels of questions in assignments and assessments help to gain a better understanding of how students comprehend to in turn complete tasks.

Alignment refers to how well all policy elements in a system work together to guide instruction and, ultimately student learning. Of the many different types of validity, (Messick, 1989, 1994; as cited by Webb, 1997) Norman Webb has revitalized and brought awareness to the analysis of assessments and curriculum. According to Webb (1997), alignment of assessments with expectations can improve the efficiency and effectiveness, which in turn promotes student learning and information attainment (p. 9). With the increased importance imposed by the Every Student Succeeds Act (ESSA), procedures for determining the alignment assessments and standards have gained the attention of our educational stakeholders most significantly political individuals that require assessments to meet a criteria of alignment. According to Webb (2007), the criteria of such are alignment should include comprehensiveness, content and performance match, emphasis, depth, consistency with performance standards, and clarity for users. Dr. Norman Webb has forced states to rethink the content assessed and its intended cognitive demand. 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, et al., 2009). DOK consistency between standards and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards (Webb, 2009). The objectives within the standards and assessment should be comparable, and according to Webb, 1997,1999, 2007, is an essential requirement of alignment analysis. Webb's DOK affords students the opportunity to articulate, gain deep understanding, and related to content. The levels of cognitive development are (Hess, et al., 2009; Webb, 1997):

- DOK-1 (Recall & Reproduction)- Recall a fact, term, principle, or concept; perform a routine procedure.
- DOK-2 (Basic Application of Skills/Concepts)- Use information, conceptual knowledge; select appropriate procedures for a task; perform two or more steps with decision points along the way; solve routine problems; organize or display; data; interpret or use simple graphs.
- DOK-3 (Strategic Thinking)- Reason or develop a plan to approach a problem; employ some decision-making and justification; solve abstract, complex, or non-routine problems, complex. (DOK-3 problems often allow more than one possible answer.)

- DOK-4 (Extended Thinking)- Perform investigations or apply concepts and skills to the real world that require time to research, problem solve, and process multiple conditions of the problem or task; perform non-routine manipulations across disciplines, content areas, or multiple sources.

Webb (2002, 2008) conducted studies regarding the alignment analysis conducted on the standards and assessments of three states and reading standards and assessments in grades 3-8 and 10. The analysis of both studies judged the alignment between the standards and the assessment using four criteria (Webb, 2002). The first criterion categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents (Webb, 2002, p.3). Webb aligned the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD), a standardized test, and the state standards for reading and mathematics. (Webb 1997 as cited by Sydoruk, 2018) In both studies, Webb found that there were issues with categorical concurrence, in that not all items were aligned between the assessment and the supposed corresponding standards (Sydoruk, 2018).

The second criteria depth of knowledge examines the alignment between the standards, assessment and if in fact cognitively demanding as what students are expected to learn and perform in their grade level (Webb, 2008). One of the studies employed the six stages Extended Depth of Knowledge Stages for Special Education (EDOK) instead of the traditional four depth-of-knowledge levels. Interpreting and assigning depth-of-knowledge levels to both objectives within standards and assessment items is an essential requirement of alignment analysis. EDOK partitions the first DOK level (Recall and Recognition) into three stages, respond, reproduce, and recall (Webb, 2008). At Stage 1, students are expected to respond or acknowledge text, such as pointing to letters or words or providing a response to conversation. Students must be able to copy or replicate in Stage 2. Stage 3 expects students to be able to recite or recall information, such as the identification of pictures, letters, and details in text.

According to the depth of knowledge consistency criterion, both studies yielded similar results. In the first study Webb (2002) determined that nearly all of the standards and assessments analyzed failed to fully meet an acceptable level on the depth of knowledge criterion. The second study also identified issues with alignments to the criterion, in that at least one item was not properly met for at least one standard, existing in Grades 3,5, 6,7, 8, and 10 (Webb, 1997 as cited by Sydoruk, 2018).

The third criterion, the range-of-knowledge is used to judge whether the knowledge expected is comparable to the span of knowledge that students need in order to complete assessment/ activities (Webb, 2008). Webb found that generally all standards met this criterion. Lastly, Webb identifies a criterion known as balance and representation, which compares the emphasis given to a particular objective on an assessment compared to other objectives, aiming to ensure a balance between each objective being assessed (Webb as cited by Sydoruk, 2019). The alignment analysis of extended standards and assessments conducted by Webb (2008) yielded acceptable results, but in the second study the results were medial (Webb, 2002).

Burns (2017) sought to use the DOK levels to describe and compare the percentages of the New Jersey Student Learning Standards and of the former New Jersey Core Curriculum Content Standards in Grades 6–8 mathematics that required students to demonstrate strategic and/or creative thinking.

Hess' Cognitive Rigor Matrix.

The CRM superimposes Blooms Taxonomy six levels and Webb's DOK. Hess' Cognitive Rigor Matrix developed in 2005, combined two models for describing the complexity and deeper learning that were accepted in the fields of education and assessment (Hess, et al., 2009b). Bloom's Taxonomy and Webb's DOK differ in scope and application. According to Hess, "Bloom's Taxonomy categorizes the cognitive skills required of the brain to perform a task, describing the type of thinking necessary to answer a question, and the Depth of knowledge, relates more closely to the depth of content understanding and the scope of a learning, which manifests in the skills required to complete the task from beginning to end (e.g., planning, researching, drawing conclusions)" (Hess, et al. 2009a, p. 3). The CRM allows educators to examine the depth of understanding required for the performance of different tasks that might seem comparable to the levels of complexity (Hess, et al., 2009b). Below is Hess's Cognitive Rigor Matrix with specific English Language Arts and Social Studies examples:

Applying Webb's DOK Levels to Bloom's Taxonomy of Educational Objectives (Karin Hess)

ELA/Soc St Examples	Webb's Depth of Knowledge Levels			
Bloom's Taxonomy	**Level 1** **Recall & Reproduction**	**Level 2** **Skills & Concepts**	**Level 3** **Strategic Thinking/ Reasoning**	**Level 4** **Extended Thinking**
Knowledge Define, duplicate, label, list, memorize, name, order, recognize, relate, recall, reproduce, state	o List/generate ideas for writing or research o Recall, recognize, or locate basic facts, ideas, principles, concepts o Identify/describe key figures, places, or events in a particular context			
Comprehension Classify, describe, discuss, explain, express, identify, indicate, locate, recognize, report, restate, review, select, translate	o Write a simple sentence o Select appropriate word(s) to use in context when meaning is evident o Identify or describe characters, setting, plot, problem, solution o Describe or explain: who, what, where, when	o Determine or recognize main idea/generalizations o Take and organize notes around common ideas/topics o summarize ideas/events o Make basic inferences or logical predictions from text o Explain relationships/cause-effect	o Write full composition using varied sentence types & structures to meet purposes o Explain, generalize, or connect ideas using supporting evidence o Make inferences about theme or author's purpose	o Write full composition demonstrating synthesis & analysis of complex ideas o Compare multiple works by same author, across time periods, genres, etc.
Application Apply, choose, demonstrate, dramatize, employ, illustrate, interpret, practice, schedule, sketch, solve, use, write	o Apply spelling, grammar, punctuation, conventions rules in writing o Use structures (pre/suffix) or relationships (synonym) to determine word meaning o Use resources to edit/revise	o Write paragraph using a basic structure or template o Edit final draft for mechanics and conventions o Use context clues to determine meaning o Use text features to find information	o Edit final draft for meaning/progression of ideas o Apply a concept in other/new contexts o Support ideas with examples, citations, details, elaboration, quotations, text references	o Define and illustrate common social, historical, economic, or geographical themes and how they interrelate
Analysis Analyze, appraise, calculate, categorize, compare, criticize, discriminate, distinguish, examine, experiment	o Identify specific information contained in maps, charts, tables, graphs, or diagrams	o Analyze a paragraph for simple organizational structure o Determine fiction/ nonfiction; fact/opinion o Describe purpose of text features o Identify use of literary devices	o Analyze an essay o Compare information within or across text passages o Analyze interrelationships among text elements, situations, events, or ideas o Analyze use of literary devices	o Analyze multiple works by the same author, across time periods, genres, o Analyze complex/abstract themes
Synthesis Rearrange, assemble, collect, compose, create, design, develop, formulate, manage, organize, plan, propose, set up, write	o Brainstorm ideas, concepts, or perspectives related to a topic		o Synthesize information within one source or text o Develop a model for a complex situation	o Synthesize information across multiple sources or texts o Given a situation/problem, research, define, and describe the situation/problem and provide alternative solutions
Evaluation Appraise, argue, assess, choose, compare, defend estimate, judge, predict, rate, select, support, value			o Cite evidence and develop a logical argument for concepts o Make & support generalizations, using text evidence	o Gather, analyze, & evaluate information to draw conclusions o Evaluate relevancy, accuracy, completeness of information from multiple sources

Figure 4. ELA/ social studies DOK levels to Blooms Taxonomy of educational objectives (Karin Hess) (Hess, 2006).

The Cognitive rigor matrix was used in an analysis of two large-scale studies of mathematics and English language arts curricula; teachers from 200 Nevada and Oklahoma public schools submitted 200,000 work samples. (The Standards Company LLC, 2008a, 2008b as cited by Hess, et al, 2009a). Curriculum specialists analyzed the items on work samples using the CRM, assigning to each sample its DOK level and the highest Bloom's Taxonomy level appearing on the sample (Hess, et al, 2009a). Results for English language arts indicate a preponderance of assignments correlating to the [2, 2] cell of cognitive rigor. (The two coordinates denote the levels of DOK and Bloom's Taxonomy, respectively.) mathematics assignments, on the other hand, heavily sampled the [1,1] and [1,3] cells (Hess et al., 2009). The tool affords educators the opportunity to properly analyze curriculum and assessment for cognitive rigor; they can then provide students with cognitively appropriate instruction that prepares them for global competitiveness.

The Partnership for Assessment for Readiness for College and Careers

Common Core State Standards (CCSS, 2010) claimed the standards were designed to provide and build upon the most advanced critical thinking, problem solving, and analytical skills that will in turn prepare students for success in college, career, and life (CCSS, 2010). The standards are measured through standardized assessments developed by the Partnership for Assessment of Readiness for College and Careers. PARCC is a consortium of states that has developed next-generation assessment system in English and math anchored in what it takes to be ready for college and careers (Camara & Quenemoen, 2012). The PARCC assessments are summative (Brown, Afflerbach, & Croninger, 2014). 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 difficulty 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

PARCC's (2019) early and continuing design commitments reflect the Partnership's ambitions to meet these high expectations for next-generation, college and career readiness assessments. 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). Herman & Linn (2014), summarizes PARCC and Smarter Balanced claims in English language arts:

- Reading: Students can independently read and closely analyze a range of increasingly complex texts.
- Writing: Students can produce well-grounded and effective writing for a variety of purposes and audiences.

- Research: Students can build and present knowledge through research and the integration, comparison, and synthesis of ideas. Likewise, here's a summary of PARCC and Smarter Balanced claims in mathematics:
- Concepts and Procedures: Students can explain and apply mathematical concepts and procedures and carry out mathematical procedures with precision and fluency.
- Problem Solving: Students can solve a range of complex, well- posed problems in pure and applied mathematics.
- Communicating/Reasoning: Students can clearly and precisely construct viable arguments.
- Modeling and Data Analysis: Students can analyze complex, real world scenarios and construct and use mathematical models to interpret and solve problems.

According to Herman & Lin (2013), results from previous studies indicate that PARCC and Smarter Balanced summative assessments are likely to represent important goals for deeper learning, particularly those related to mastering and being able to apply core academic content and cognitive strategies related to complex thinking, communication, and problem solving. To support the claims about assessments' complexity the PARCC consortium developed and employed the PARCC cognitive complexity frameworks. The consortiums master claim in terms of mathematics is keeping students on-track for college and career readiness. To achieve this the student must solve grade-level /course-level problems in mathematics as set forth in the Standards for Mathematical Content with connections to the Standards for Mathematical Practice (PARCC, 2019).

In the study *Measuring Deeper Learning through Cognitively Demanding Test Items: Results from the Analysis of Six National and International Exams*, Yuan & Le (2014) described how PARCC uses two frameworks for mathematics and ELA/literacy and how it is defined in terms of sources of cognitive complexity. Mathematics employs five sources of cognitive complexity. The mathematical content, in each grade level impose a demand of complexity, and each level is categorized from a range of low to high complexity. The source of mathematical practices involves how students are expected to perform and how it is applied. Stimulus material accounts for the role of technology and response mode examines the requirements in which a student must complete the assessment. Lastly, processing demands explains the reading and linguistics demand in each item. Below is the Sources of Cognitive Complexity in Mathematics:

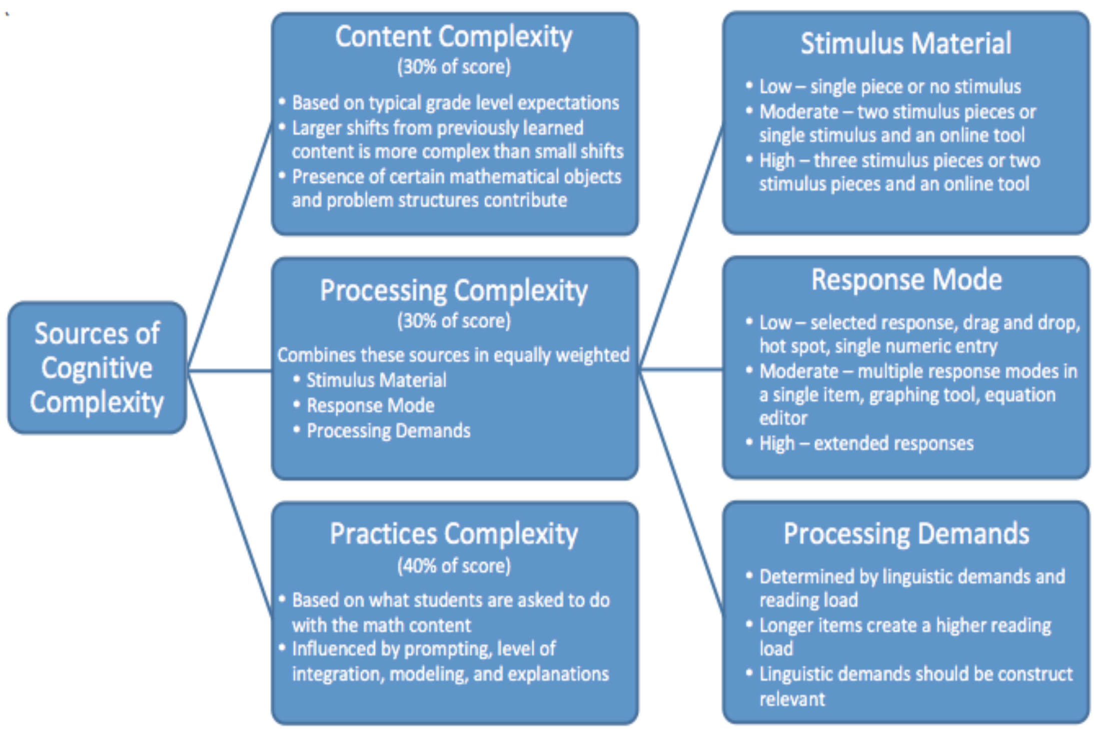

Figure 5. Proposed sources of cognitive complexity in PARCC items and tasks: mathematics (PARCC, 2019).

In the Summative ELA/literacy assessments, students are expected to read and analyze fiction and nonfiction passages. They must also be able to write what they learn by using evidence to support their arguments. In the English language arts (ELA)/literacy assessments, the PLDs (performance level descriptors) at each grade level are written for the two assessment claims of reading and writing (PARCC, 2019). PLDs indicate what a student in each grade level must demonstrate. The reading claims are text complexity, range of accuracy in reading comprehension and responses, and the evidence cited. For the writing claims, PLDs are differentiated in the two factors written expression, and knowledge of language and conventions.

To support both the reading and writing claims, PARCC employs four sources of cognitive complexity to analyze items and tasks (Yuan & Le, 2014). Text complexity, a text will be assigned to one of the three categories of complexity (readily accessible, moderately complex, or very complex). The source of command of textual evidence defines the amount of text that a student must process and understand. As mentioned previously, the response mode is the way students are expected to answer to complete the assessment and can consequently influence the items cognitive demand. Lastly, the processing demands explain the affects in cognitive complexity within the linguistic demands and reading. The PARCC CCR (College and Career Ready) Determinations in ELA/Literacy and mathematics describe the academic knowledge, skills, and practices that students must demonstrate to show their ability to directly enter and succeed in entry-level and relevant technical courses in content areas at two- and four-year public institutions of higher education (PARCC, 2019). Below you can find the Sources of Cognitive Complexity in ELA/Literacy used for assessments determinations.

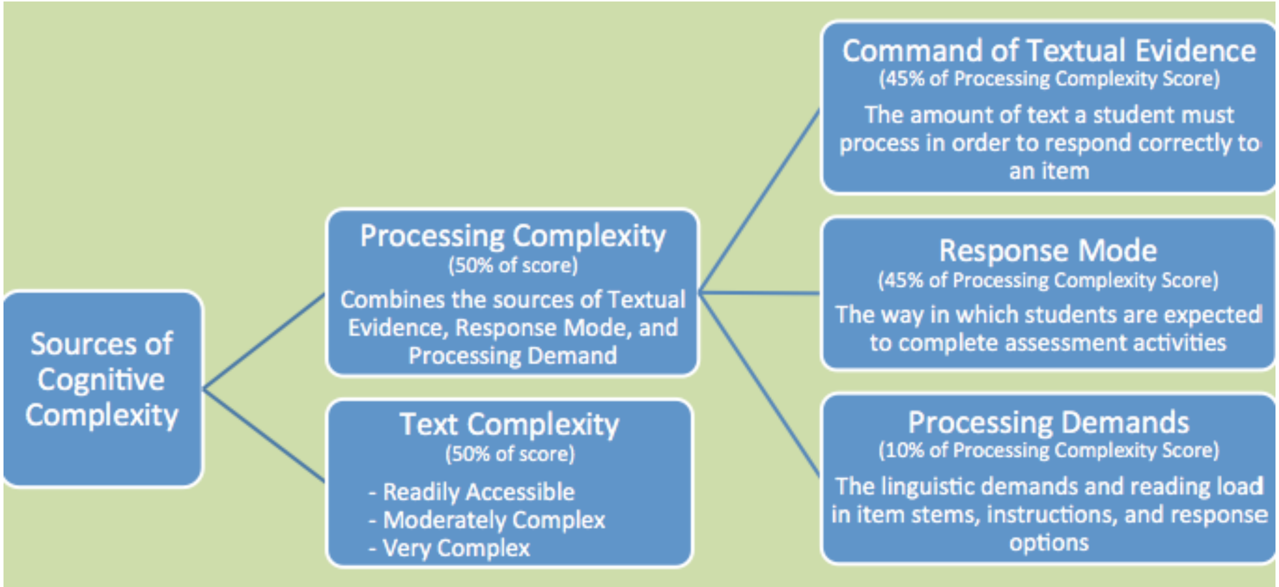

Figure 6. Proposed sources of cognitive complexity in PARCC items and Tasks: ELA/literacy (PARCC, 2019).

The use of results from standardized assessments to make important judgments about students and teachers have been considered a controversial aspect of American education. Since the passage of the No Child Left Behind Act more then 15 years ago, state tests have played an outsized role in schools (Rothman & Marion, 2016). The law required states to develop high-quality academic assessments aligned with challenging state academic standards that measure students' knowledge of reading/language arts, mathematics, and science (Ashby, 2009). The overarching goal for this law was to incentivize educators to focus on the student's achievement. PARCC's critics have long complained that the exams are not grade appropriate, which is another way of saying they're unfairly challenging (Plotting future without PARCC). Tying these results to evaluations have caused pressure to "teach to the test".

PARCC has been dropped by many states, with the illusion of new assessments. In 2017, the PARCC consortium sold its test questions to the Council of Chief State School Officers, which represents state education commissioners (Gerwertz, 2019). Moreover, PARCC chose a new organization, New Meridian, to manage licensing those test questions to states (p.2). In the article *Parking PARCC Claims in the Dumpster of Failed Reforms* (2019), Tienken, discussed the fading of PARCC, and how its questions and similar tests are developed through the mentioned licensing agreements in 10 states and the New Meridian Corporation. Originally the consortium had an all or nothing mentality, but with the exit of states, the usage of the item bank (test questions) was then implemented as a survival tactic to keep them afloat.

Tienken (2019) argued that the four common claims made to the public by state education agency personnel and PARCC officials. The following claims and arguments are:

Claim 1 (*PARCC tests are diagnostic and the results provide educators with important information about student mastery*): Assessments must have reliability figures around 0.80 to 0.90 to diagnose a student’s achievement of any skill at the individual level. To attain reliability the test must include about 20 to 25 questions per skill (Frisbie; Tanner, as cited by Tienken, 2019). Furthermore, the PARCC assessment tends to assess multiple skills in one question. The information the teacher may gain to further help the student is received after they have subsequently transitioned to the next grade level. We can infer that the assessment does not provide an accurate glimpse of the student’s achievement.

Claim 2 (*PARCC provides valid results of what students know and can do in English Language Arts and Mathematics*): Tienken and colleagues completed a study regarding the predictability of the results from the PARCC Algebra 1 test and the PARCC English 10 tests administered in New Jersey during the 2016-2017 school year. The results from the study suggest that the percentage of students scoring proficient or above on the Algebra 1 and English 10 can be predicted for 71%-75% of the 159 school districts in the sample. If standardized assessments can only be predicted with a moderate level of accuracy, how valid are the results for making decisions that determine the way we educate our students?

Claim 3 (*PARCC results tell parents, students, teachers, and the public whether students in Grades 3–8 and high school are college and career ready*): If PARCC provides valuable information of whether students are college and caree ready, why isn't use as a determination instead of the SAT or ACTs? Not even the SATs can predict accurately which students will do well their first year of college or beyond (College Board, as cited by Tienken, 2019).

Claim 4 (*PARCC assesses important 21st-century skills and knowledge*): PARCC tests mostly measures 19th-century skills with a 20th-century tool. The assessments are aligned to the Common core state standards, which have the expectation of analyzing, but if looked closely students are only required to analyze for one right answer. (Tienken, 2019, pp. 57-59)

According to Tienken (2019), the ultimate assessment system already exists in public school classrooms: the teacher.

The PARCC tests have long been criticized for being administered in high-stakes circumstances before they were studied and validated. In the article *Alice in PARCCland: Does' validity study' really prove the Common Core test is valid?* (2016), William Mathis states that PARCC's rejoinder is that they had content validity, meaning that the test was built according to their committee-reviewed specifications. But what is missing is predictive validity meaning, the test results do not equate to the measure of "College and Career Ready?" (p.2) This claim is extensively emphasized in PARCC marketing materials and parroted by some education policy makers and educators.

In the study tiled *Predictive Validity of MCAS and PARCC: Comparing 10th Grade MCAS Tests to PARCC Integrated Math II, Algebra II, and 10th Grade English Language Arts Tests* (2015), the state of Massachusetts was deciding whether to continue using the MCAS (Massachusetts Comprehensive Assessment system) or adopt PARCC. Since at the time there was no research regarding PARCC and its predictability of college readiness, the Massachusetts Executive Office of education commissioned Mathematica Policy Research to conduct a study. The study was set to provide evidence on the extent to which MCAS and PARCC test scores can accurately assess whether students will succeed in college (Nichols-Barrer, Place, Dillon, & Gill, 2015).

Ultimately, it was determined by the authors of the study that the PARCC and MCAS 10th-grade exams equally predicted college success, as measured by first-year grades (GPA) and probable that the students would need remediation (Nichols-Barrer et al., 2015). Employing correlation coefficient, a sample size of about 847 college freshmen was divided into two MCAS testing groups and five PARCC testing groups. Correlation coefficient is a statistical measure of the relationship between test scores and college outcomes (Nichols-Barrer et al., 2015). The correlation coefficient provides a common benchmark to summarize the relationship between two variables.

Mathematica determined that the correlation coefficients between test scores and GPA were, ranging from 0.07 to 0.40 (p.11). The correlations between math GPA and PARCC math scores are 0.37 to 0.40. The ELA test and ELA GPA had a small correlation of 0.13 to 0.26 (p.11). There are a number of issues associated with the use of correlation which includes the effect of non-linear relationships, outliers, restriction of range, correlation versus causality and statistical versus practical significance (Pallant, 2010). The study sample only included college students instead of real high school students. Correlation coefficients in this study ran from minus one (perfect inverse relationship) to zero (no relationship) to 1.0 (perfect relationship) (Mathis, 2016). How much one measure predicts another is the square of the correlation coefficient (p.4). For instance, when you square the highest coefficient (0.40) it gives us .16, subsequently meaning that PARCC tests predicted 16 percent of first-year college GPA (Mathis, 2016). When computing the rest of the correlations one can determine that most of the sample size was not utilized. With such low predictability, Mathematica fails to provide accurate information. Its implications that standardized tests predict college readiness is unfounded, and one can further determine how a play in numbers can paint another story.

Related Studies

There are many questions regarding the validity of standardized assessments, with limited research that affirms their claims. Standardized assessments evaluate what the students are expected to gain through curriculum and instruction. The following studies systematically examine to what extent is higher learning embedded in the assessments and programs using cognitive complexity tools.

Six National and International Exams

In 2010, the William and Flora Hewlett Foundation's Education Program launched its strategic Deeper Learning Initiative, which focuses on students' development of deeper learning skills (i.e., the mastery of core academic content, critical-thinking, problem-solving, collaboration, communication, and "learn-how-to-learn" skills) (Yuan & Le, 2014, p.xi). In the study Yuan and Le (2012b), examined the cognitive demand of six nationally and internationally administered tests with the goal of providing a benchmark of understanding to the extent these large-scale assessments—and, measure students' deeper learning.

The study employed two frameworks Webb's Depth of Knowledge and the PARCC framework to analyze three deeper learning skills: critical thinking, problem, solving, and written communication (Yuan and Le 2012b). Webb's DOK defines four levels of cognitive complexity. PARCC provides two frameworks to describe the cognitive demands of mathematics and ELA (Yuan and Le, p.xii). Although the PARCC framework provided guidelines for combining various dimensions to create an overall complexity score, Yuan and Le (2012b) deviated from the recommended scoring mechanism. It appeared that the DOK framework placed relatively greater emphasis on the types of cognitive processes elicited, whereas the PARCC framework placed relatively greater emphasis on the difficulty of the content being tested (Yuan & Le, 2014).

The six assessments varied in their results regarding the cognitive demands. IB and AP had higher percentages of cognitively demanding items than other benchmark tests in both subjects compared to TIMSS and PIRLS, which appeared to be less cognitively demanding than other benchmark tests (Yuan & Le, 2012b). There was indication that the percentage of cognitively demanding items on the six tests was associated with the purpose of the assessments and the targeted student population (Yuan & Le, 2012b). The IB and AP tests assess students' readiness for postsecondary academic learning and target academically advanced high school students, in contrast, PISA, NAEP, TIMSS, and PIRLS assess what students know and can do at the time of the administered test (p.15).

SMARTER Balanced Assessment Consortium Common Core State Standards

The Smarter Balanced Consortium like PARCC developed a standardized assessment aligned to the Common Core State Standards. Essentially, the two consortiums seek to meet the same goal, determining if a student is “college and career ready”. Both assessments utilize computer-based ELA and math assessments designed to evaluate proficiency in Grades 3-8 and high school.

Sato, Lagunoff, and Worth’s (2011) study was a descriptive analysis of the Common Core State Standards, determining which content is eligible for the Smarter Balanced Assessment (SBA) Consortium’s end of year summative assessment for ELA and mathematics in grades 3-8 and high school. The high school standards analyzed were those in grades 9-10 and 11-12 ELA, and all conceptual categories for mathematics (Sato et al., 2011).

Sato et al.’s (2011) analysis aimed to address two key questions: which CCSS are eligible for the SBAC summative assessment, and the range of depth of knowledge. In order to determine eligibility, content standards were coded according to the criteria and coding dimensions (i.e., learnable during the school year, expected of all students, measurable via on-demand assessment, eligible for the summative assessment, response type, and DOK) (Sato et al., 2011). Employing Webb’s DOK, coders reviewed each standard to determine the range of cognitive complexity required to preform the skill or demonstrate the knowledge described by the standard (Sato et al., 2011). Subsequently, the findings from the DOK coding provide information on the range of cognitive complexity of content in the standards, which helps with the development of the assessment.

The findings in this analysis were intended as a starting point in the development of the standardized assessment. In ELA Sato et al. (2011) employed the pattern for DOK levels, and determined its similarities is for all standards and eligible standards. Across all grades, the majority of standards were coded to DOK Levels 2 and 3, with the number coded to DOK Level 2 decreasing slightly and the number coded to DOK 3 increasing slightly from the elementary grades to the secondary grades. Standards coded to DOK Level 4 increased from grades 3 through 6, and became constant between grades 7 and 9–10, rising slightly at grades 11–12. Standards coded to DOK Level 1 followed the reverse pattern, decreasing from grades 3 through 5, and remaining about the same at grades 6 through 12 (p.19). In mathematics across all grades and conceptual categories, the majority of standards were coded to DOK Level 1 and/or Level 2. In grade 7, grade 8, and especially the high school conceptual category Geometry, a notable number of standards were also coded to Level 3 (Sato et al., 2011). One standard in Geometry was coded to Level 4 (p.34).

Higher Order Thinking Requirements of an Online-Based ELA Skills Program

Online-based programs have gained popularity in recent years, there is little research conducted on the effectiveness of these programs, the validity of the claims made by these private companies, or the types and frequency of tasks that promote higher order thinking skill set development embedded in such programs (Sydoruk, 2018). This program, as well as many others, was designed to meet the needs of 21st century skills, which had been incorporated into Standard 9 of the NJCCCS (p.106). Sydoruk (2018) employed the Hess Cognitive Framework to analyze the level and distribution of cognitive complexity within HOT Learning program.

Sydoruk explored the topic by analyzing Grade 8 English language arts questions from the program using CRM as an analytical framework to categorize the distribution of higher order thinking of a question (Sydoruk, 2018). Two coders utilized the framework to examine 231 questions from the HOT Learning program following the double-rater read-behind consensus model (Sydoruk, 2018).

The findings in this study yielded that the program was not considered to be cognitively complex. Of the questions examined in the HOT Learning program, 12.1% of questions placed into higher-level cells (p.106). There were 203 questions, or 87.9%, from the sample that placed into Levels 1 and 2 (p.101). 139 questions placed into Level 2, making up 68.5% were placed into cell [2,2] required students to choose a main idea that best fit the passage following a multiple-choice format (Sydoruk, 2018). 64 questions placed into Level 1, which equaled 31.5% of were placed into cell [1,2] because they required students to describe an event that happened in the passage or to define a term from the text (p. 101).

Theoretical Framework

Numerous theoretical frameworks have been developed throughout the years, and have sought to be used by many in the United States. Specifically, Blooms Taxonomy and Webb's Depth of Knowledge have gain popularity due to its use for the analyzing cognitive demand. The frameworks have similarities regarding the analysis of higher order thinking, but also have differences in its applications. Blooms Taxonomy specifically focuses on the action, measuring students' abilities and outcomes according to the six cognitive levels. The revised framework moved away from nouns and places focus on verbs to facilitate the action of higher order thinking.

On the contrary, the Webb (1997, 2002) alignment process is one of a handful of processes that have been used to determine cognitive demands between curriculum standards and assessments (Blank, 2002). Many states and districts employ DOK to designate the depth and complexity of state standards to align the state's large- scale assessments or to revise existing standards to achieve higher cognitive levels for instruction (Hess et al., 2009a). The language arts and mathematics CRM tables found in Appendix A and Appendix B illustrate the DOK levels employed to analyze each practice test question. As mentioned previously educational stakeholders have agreed in the importance of defining higher order thinking and its application.

This study utilizes Hess' Cognitive Rigor Matrix as the framework to categorize the complexity of language used on PARCC practice test questions (Hess et al., 2009a). The CRM superimposes two different cognitive complexity measures – Bloom's Taxonomy and Webb's Depth of Knowledge – to produce a means of analyzing the emphasis placed on curricular materials, instructional focus, and classroom assessment (Hess, 2013). Utilizing 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 draw important conclusions.

PARCC claims to provide all students with equitable access to high-quality, 21st century assessments (PARCC, 2018). The theoretical framework aims to compare the practices tests from Grades 3 and 4 English language arts and mathematics with the PARCC claims, while describing the level and distribution of higher order thinking. A major obstacle when assessing the complexity of the assessments questions is the non-definitive definitions of higher order thinking. Despite a plethora of research that highlight the actions and tasks that contribute to higher order thinking, such as critical thinking and problem solving, there is no unified definition of higher order thinking to which educators and researchers can refer (Sydoruk, 2019). Educators tend to believe that they're applying higher order thinking into their lessons and assessments when they are not. Studies analyzing classroom tests, over many decades, have found that most teacher-made tests require only recall of information (Marso & Pigge, 1993 as cited by Brookhart, 2010).

When teachers are surveyed about how often they think they assess application, reasoning, and higher-order thinking, teachers claim they assess these cognitive levels quite a bit

(Brookhart, p.10). There is an array of strategies that educators can implement to produce higher order information from their learners. The application of higher order thinking is developed utilizing various learning activities and assessments including critical thinking, problem solving, reasoning, and creative thinking (Brookhart, p.14). Hess' Cognitive Rigor Matrix provides a mean of analyzing the higher order thinking within curricular activities and assessments.

Hess' Cognitive Rigor Matrix (CRM) will be employed to analyze complex thinking in a nationally used standardized test. This study ultimately aims to systematically examine the extent in which higher learning is embedded in the PARCC practice elementary tests. The essential goal is to gather and provide information that will help create a new education paradigm that will cultivate creative and entrepreneurial talents to make creativity, entrepreneurship education the core to the education (Zhao, 2012). Hence, 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: Methodology

Introduction

The purpose for this Mixed-methods study was to compare, analyze, and describe the language of complex thinking embedded within the 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in language arts and mathematics Grades 3 and 4. Policy focus on both test-driven accountability and 21st century skills (sometimes called higher order thinking, complex thinking, deeper learning etc.) is accelerating (Nehring, Charner-Laird, M, & Szczesiul, 2019). Employers, postsecondary institutions, and civic leaders are urging greater focus on 21st century skills essential for college, career, and civic success: problem solving, interpersonal skills, and collaboration (Parsi & Darling-Hammond, 2015). In response to these demands, states across the United States are working to readjust policies on educational standards, standardized assessments, and human capital strategies to set a new course for their state education systems (Nehring et al., 2019; Parisi & Darling, 2015; Every Student Succeeds Act, 2016; No Child Left Behind, 2002).

The Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests 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. Although the name PARCC might be disappearing into the past, PARCC questions and PARCC-like tests will live on through shared licensing agreements between states and entities (Tienken, 2019). 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 2016 PARCC practice tests in English language arts and mathematics in grades 3 and 4?

1. In what way(s) does the language of the questions on the English language arts section of 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 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 mathematics section of 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 associate with the language that promotes higher-order thinking found in research literature?
3. What is the distribution of thinking on the 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in English language arts and mathematics in Grades 3 and 4?

Policy Context

In 2010, states joined together to develop and adopt the Common Core State Standards (CCSS) with the intent of preparing the students for success in college and career (PARCC, 2019). According to the PARCC consortium (2019), the assessment is based on research and benchmarking to the standards of high performing nations, and to the demands of rigorous college courses (PARCC, 2019). PARCC's claims directly reflect a policy concern about measuring the 21st century cognitive competencies (PARCC, 2019).

The PARCC consortium had many participant states, although some of those dropped out of the project before the first test administration, in the 2014-15 school year. Under parent and educator backlash against PARCC, the states that dropped the standardized assessment, decided to replace the exam. New Mexico and Maryland being the latest states to indicate an end to PARCC testing (Tienken, 2019). As mentioned previously, the PARCC assessment may be fading, but PARCC questions and PARCC-like tests will live on (Tienken, 2019). An example would be states such as Illinois, whom do not intend on cutting the cord with PARCC, instead includes a core of PARCC test items on the new tests so that it can maintain some level of year-to-year comparability in student results (Sawchuk, 2018). In essence, the items in the assessment are similar or a complete copy of the PARCC questions.

The consortia's goal was to put forth an assessment that could assess student preparedness for college and career. According to Herman and Linn (2014), the United States invested in the PARCC consortia to develop assessment systems that would embody the Common Core State Standards (Herman & Linn. 2014). The PARCC College and Career Ready (CCR) Determinations in ELA/Literacy and mathematics describes the academic knowledge, skills, and practices in English language arts/literacy and mathematics students must demonstrate to show they are able to enter directly into and succeed in entry-level, credit-bearing courses and relevant technical courses in those content areas at two- and four-year public institutions of higher education (PARCC, 2019). Subsequently, the consortia claimed that CCR Determination would provide policymakers, educators, parents, and students with a clear signal about the level of academic preparation needed for success (PARCC, 2019).

Research Design

The design for this study was a case study with Mixed-methods. "A case study is an in-depth description and analysis of a bounded system" (Merriam & Tisdell, 2015 p. 37). Similarly, Creswell (2009) described case studies as strategies researchers explore in depth a program, event, activity, process, and or individuals (p. 13). Furthermore, Yin (2014) stated "a case study is an empirical inquiry that investigates a contemporary phenomenon its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" (Merriam & Tisdell pp. 37-38).

Most definitions regarding case studies are parallel in their beliefs. Although Merriam's (2015) definition of a qualitative case study is that of an in-depth description and analysis of a bounded system, it is congruent with other definitions (Bogdan & Biklen, 2007; Cresswell, 2007; Patton, 2002; Stake, 2005). Freebody (2003) explained how case studies focus on one particular instance of educational experience and attempt to gain theoretical and professional insights from a full documentation of that instance.

The case study design was employed in this study because it provided the structure and methods needed to study the cognitive complexity within the PARCC assessments. Additionally, the design afforded the opportunity to put in place an inquiry in which both researchers and educators can reflect upon particular instances of educational practice (Freebody, 2003 p.103).

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 3 and 4 based on pre-existing codes. Qualitative content analysis is a research approach for the description and interpretation of textual data using the systematic process of coding (Assarroudi, Nabavi, Armat, Ebadi, & Vaismoradi, 2018). Creswell (2009) described qualitative data analysis as an ongoing process involving preparing the data for analysis, conducting different analyses, moving deeper into understanding the data, representing the data, and making an interpretation of the larger meaning of the data (p.183).

The final product of data analysis is the identification of categories, themes and patterns (Elo and Kynga¨ s, 2008; Hsieh and Shannon, 2005; Zhang and Wildemuth, 2009 as cited by Assarroudi et al., 2018). Hsieh and Shannon (2005) emphasized that the success of a content analysis depends greatly on the coding process. Creswell (2009) defined coding as the process of organizing the material into chunks or segments of text before bringing meaning to information (Rossman & Tallis, 1998).

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 Grades 3 and 4 PARCC practice assessments in English language arts and mathematics based on the Hess Cognitive Rigor Matrix methodology (See Figure 5). 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 220 English language arts and mathematics questions obtained from the PARCC practice tests. "In deductive content analysis, the organization phase involves categorization matrix development, whereby all the data are reviewed for content and coded for correspondence to or exemplification of the identified categories" (Polit & Beck, 2012 as cited by Elo et al., 2014, p.2). Deductive category application works with prior formulated, theoretical derived aspects of analysis, bringing them in connection with the text (Mayring, 2000). The language on Hess' Cognitive Rigor Matrix was used to categorize the PARCC questions according to its complexity. The coding and analysis process is outlined in the figure below adapted from Mayring (2000).

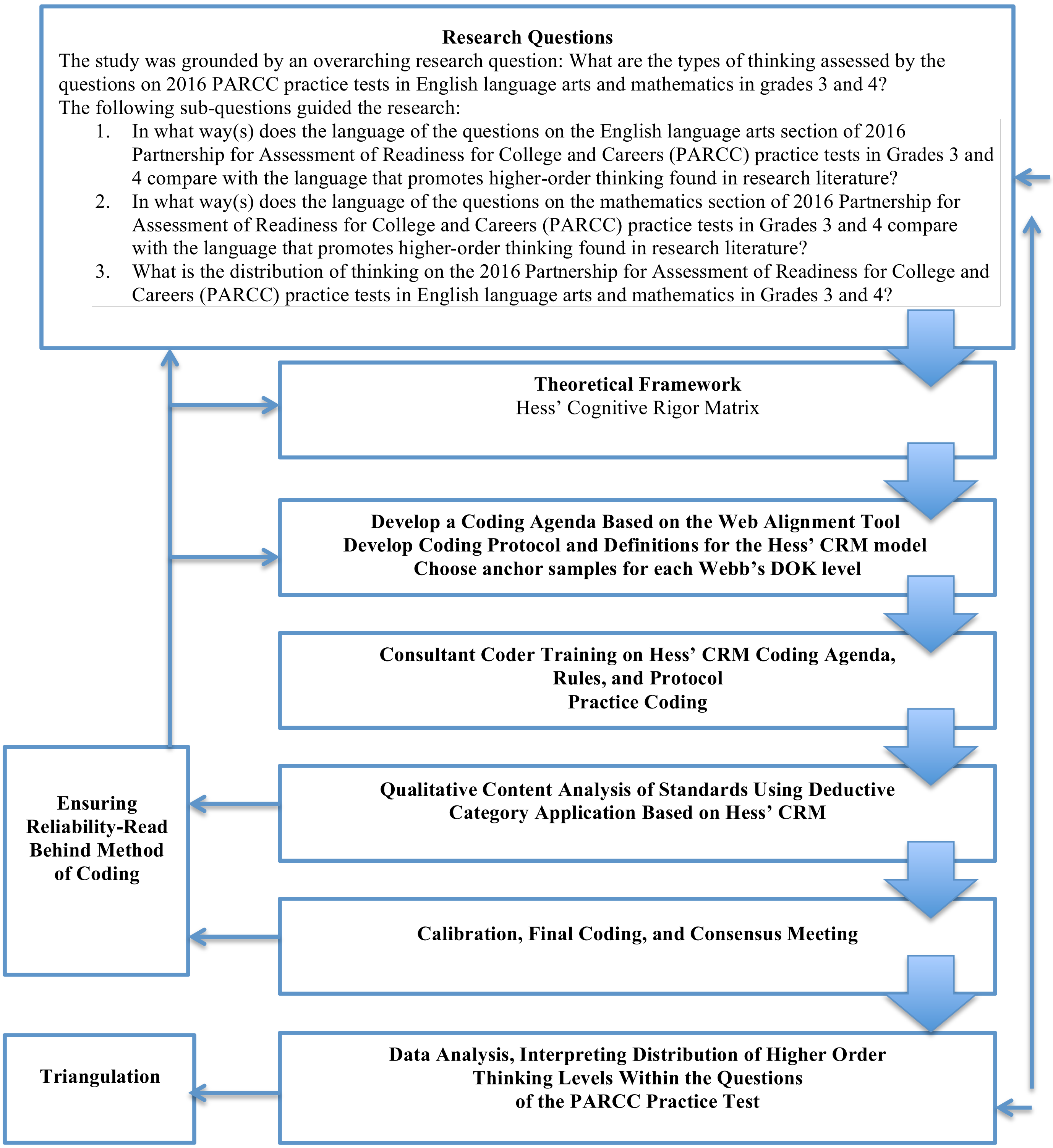

Figure 7. Step model for deductive category application, adapted from Mayring (2000).

After assessing the authenticity and nature of documents or artifacts, the researcher must adopt some system for coding and cataloging them (Merriam, p.152). In this study, the PARCC assessment questions Grades 3 and 4 in language arts and mathematics were coded and analyzed based on Hess' CRM. Hess' Cognitive Rigor Matrix was appropriate for this study because the language of its framework included much of the language of higher order thinking found in the literature. The Matrix was an organized way to categorize the types of complex thinking in the test questions.

According to Hess et al. (2009a), the Cognitive Rigor Matrix (CRM) 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. Furthermore, the CRM combines the higher order thinking of Webb's Depth of knowledge and the analysis of cognitive skills within tasks and assessments. "The resulting combination of Bloom's Taxonomy and depth of knowledge, cognitive rigor forms a comprehensive structure for defining rigor, thus posing a wide range of uses at all levels of curriculum development and delivery" (Hess et al., 2009a).

Quantitative methods were employed in the second part of the study. Specifically, frequencies, and descriptive statistics were utilized to describe the differences and similarities of complex thinking that exist in the language of the 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.

Description of Documents

Document is often used as an umbrella term to refer to a wide range of written, visual, digital, and physical material relevant to the study (including visual images) (Merriam & Tisdell, 2015). Documents, as the term is used in (Merriam 2009), also include what LeCompte and Preissle (1993) defined as artifacts "symbolic materials such as writing and signs and non-symbolic materials such as tools and furnishings" (p. 216). Most documents and artifacts exist prior to commencing the research study, and are produced for reasons other than the research at hand (Merriam, 2009; Merriam & Tisdell, 2015).

The English language arts and mathematics in the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 were the documents analyzed in this study. Document were downloaded from the PARCC website on July 9, 2019. The practice tests in its entirety are a 346-page document that focuses on both language arts and math subjects in grades 3 and 4. 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 were retrieved from a public website containing PARCC assessment information and various tools. The practice tests were made readily available in Grades 3-11. For this study, I only focused in Grades 3 and 4 language arts and mathematics.

Coders

As part of this study a coding committee was established. Two coders were used. The first coder has been an educator for over 10 years in Grades K-6th. 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, an educator with 8 years of experience, earned her doctorate in Educational Leadership in 2018. She has previous coding experience using the Hess CRM since 2016, and is currently using the model to code for various New Jersey schools. The coders followed and implemented the rules adapted from the Webb's Alignment Training Manual.

Coding Protocol

Hess' Cognitive Rigor Matrix is designed as a 24-cell grid, with Webb's Depth of Knowledge on the X-axis (the columns) and Bloom's Taxonomy as the Y-axis (the rows). The matrix contained cells that stipulated a specific coding scheme, in which provided how complex the document being analyzed is (Sydoruk, 2019). The Webb's Depth of knowledge being the first number and Bloom's Taxonomy being the second number for each cell (ex. [1,2]). Furthermore an explanation of each categories is provided below (adapted from Hess et al., 2009b).

- [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.
- [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, and 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.
- [2,2]: Webb's Level 2, Bloom's Level 2. 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.
- [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.
- [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.

- [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.
- [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.
- [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.
- [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.
- [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.

- [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.

The first coding practice session involved coder calibration with the two primary coders. The session was led by an expert coder who acted as the trainer. The coders agreed that questions placed into Categories 3 and 4 of Webb's DOK levels would be considered higher level, following the guidelines of the Webb Alignment Tool training manual (Webb, Alt, Ely, & Vesperman, 2005 as cited by Sydoruk, 2019). Moreover some sample rules were adapted from the Webb's Alignment Training (WAT) Manual that the coders followed when assigning Hess' level of complexity.

1. The DOK/Blooms Taxonomy levels of an objective should be the level of work students are most commonly required to perform at that grade level to successfully demonstrate their attainment of the objective.
2. The DOK/Blooms Taxonomy levels of an objective should reflect the *complexity* of the objective, rather than its *difficulty*. The DOK/Blooms Taxonomy levels

describe the kind of thinking involved in a task/assessment, not the likelihood that the task will be completed correctly.

3. In assigning DOK/Blooms Taxonomy levels to an objective, think about the complete domain of items that would be appropriate for measuring the objective. Identify the depth-of-knowledge level of the most common of these items.
4. If there is a question regarding which of two levels an objective addresses, such as Level 1 or Level 2, or Level 2 or Level 3, it is usually appropriate to select the higher of the two levels.
5. The team of reviewers should reach consensus on the DOK/Blooms Taxonomy levels for each objective before coding any items for that grade level (adapted from Webbs et al., 2005 p.38).

Additionally, the WAT was cross-referenced with Hess' CRM to include procedures for facilitating the consensus process during the formal coding process following the training session. The procedures included the following:

- Read each objective aloud before discussing it.
- As you go through the objectives, actively solicit comments from all reviewers.
- Use your printout to call on people who coded DOK/ Blooms Taxonomy levels differently from the coding of other members of the group, and ask them to explain why they coded the objective to the particular levels. Be sure they use the definitions to justify their answers.
- Once two reviewers have described how they have coded an objective differently, ask a third reviewer to highlight the differences between these two interpretations.

- Restate and summarize to reviewers your interpretation of what the reviewers have agreed on and what they have disagreed on.
- If there is a difference in interpretation of the objective's terminology or expectations, appeal to a reviewer with experience in teaching that grade level with these standards to discern how the state's teachers might be interpreting the objective.
- Ask if anyone, through other reviewers' explanations, now wants to change his or her mind about their original coding.
- If the viewpoints on the DOK/ Blooms Taxonomy levels of an objective are divided, point to the most likely skills or content knowledge required in the objective, not the more extreme possibilities the objective might allow for.
- As the facilitator, try not to dominate the consensus process. Even if you have strong feelings about the DOK/ Blooms Taxonomy levels of an objective, wait to see if other reviewers highlight your point (adapted from Webb et al., 2005).

Researcher Bias

Bias is commonly understood to be any influence that provides a distortion in the results of a study (Polit & Beck, 2014 as cited by Galdas, 2017). Recognizing and understanding research bias is crucial for determining the utility of study results and an essential aspect of evidence-based decision-making (p.1). When discussing the trustworthiness of findings from a qualitative content analysis, is important to understand that there is always a degree of interpretation when analyzing the text(Elo et al., 2014). "Thorough preparation prior to the study and data gathering, content analysis, trustworthiness discussion, and result reporting is essential" (Elo et al, p2). More than one person should perform an analysis to increase validity and to provide sound interpretation of data (Burla et al., 2008; Schreier, 2012 as cited by Elo et al., 2014). However, high inter-coder reliability (ICR) is required when more than one coder is involved (Elo et al., 2014). With the implementation of the double-rater read- behind consensus model the coding committee met, discussed and agreed the complexity of the assessment questions ensuring reliability.

Reliability and Validity

"Qualitative validity means to check for accuracy of the findings by employing procedures, while qualitative reliability indicates that the researchers approach is consistent across different researchers and different projects" (Creswell, 2009 in Gibbs, 2007 p. 201). Furthermore, Merriam and Tisdell (2015) explained the connection between reliability and internal validity from a traditional perspective rests for some on the assumption that a study is more valid if repeated observations in the same study or replications of the entire study produce the same results. This logic relies on repetition for the establishment of truth, but as everyone knows, measurements, observation, and people can be repeatedly wrong (Merriam & Tisdell, 2015).

According to Creswell (2009), the researcher could actively incorporate validity strategies that enhance the researchers ability to assess the accuracy of findings as well as convince readers of the accuracy. The following strategies can be implemented to add validity to the study (Creswell, pp.191-192):

- *Triangulate* different data sources of information by examining evidence from the sources and using it to build a coherent justification for themes. If themes are established based on converging several sources of data or perspectives from participants, then this process can be claimed as adding to the validity of the study.

- Use *member checking* to determine the accuracy of the qualitative findings through taking the final report or specific descriptions back to participants and determining whether the participants feel that they are accurate.
- Use *rich, thick description* to convey the findings.
- Clarify the *bias* the researcher brings to the study. This self-reflection creates an open and honest narrative that will resonate well with readers.
- Also present *negative or discrepant information* that runs counter to the themes. Because real life is composed of different perspectives that do not always coalesce, discussing contrary information adds to the credibility of an account. A researcher can accomplish this in discussing evidence about a theme. Most evidence will build a case for the theme; researchers can also present information that contradicts the general perspective of the theme.
- Use *peer debriefing* to enhance the accuracy of the account. This strategy involving an interpretation beyond the researcher and invested in another person adds validity to an account.
- Use an *external auditor* to review the entire project. The procedure of having an independent investigator look over many aspects of the project (e.g., accuracy of transcription, the relationship between the research questions and the data, the level of data analysis from the raw data through interpretation) enhances the overall validity of a qualitative study.

The WAT training manual provided DOK level descriptors that helped organize the complexity of tasks/ assessment (Webb et al., 2005). Webb's Alignment Tool (WAT) training manual contains definitions, explanations, and examples for coders to reference and specifically understand how the DOK levels should read for English language arts and mathematics (Sforza et al., 2016)

According to the WAT for language arts (Webb, et al., 2005):

- Level 1: Requires students to receive or recite fact or to use simple skills or abilities.
- Level 2: The engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis of inference is required. Items at this level include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion.
- Level 3: Deeper knowledge is a focal point. Students are encouraged to go beyond the text and showing understanding of the ideas presented.
- Level 4: Higher-order thinking must be present at this level. Students may be asked to develop hypotheses and perform complex analyses of the connections among texts (adapted from Webb et al., 2005).

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

- Level 4 *(Extended Thinking)*: Requires complex reasoning, planning, developing, and thinking, most likely over and 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 (adapted from Webb et al., 2005).

A double-rater read-behind consensus model was utilized to align each test question to Hess' Cognitive Rigor Matrix, which proved to be an effective strategy in previous studies (Burns, 2017; Fitzhugh, 2019; Satos et al., 2011; Sforza, 2014; Sydoruk, 2019). Specifically in Sato's et al. (2014) study, one analyst independently coded the standards; the second analyst reviewed the outcomes of the first analyst's ratings and noted agreement or disagreement with the first analyst's ratings. The analysts then discussed any discrepancies between their interpretations as necessary and reached a consensus. Through the double-rater read- behind consensus model, the coding committee held calibration sessions to discuss assessment questions. The consensus model increased inter-rater reliability and offered a means of calculating and monitoring the coders' agreement (Miles, Huberman, & Saldaña, 2014, p. 84). The consensus model was intended for descriptive purposes to inform further discussions of the assessment questions and its implications (Sato et al., 2014).

Training and Calibration

Thoroughness as a criterion of validity refers to the adequacy of the data and also depends on sound sampling and saturation (Whittemore, Chase, & Mandle, 2001 as cited by Elo et al., 2014). In order to ensure reliability and validity to the study, coders were thoroughly trained utilizing CRM (Hess et al., 2009a) and WAT training manual (Webb et al., 2005). The coding committee began training together; receiving an introduction to the goals and purpose of the study and an in-depth discussion of the study criteria, including the DOK/Blooms Taxonomy level descriptions (Sato et al., 2014). The coders discussed the specific characteristics of each category of the Cognitive Rigor Matrix and made clarifications in order to reach consensus on the meanings of the examples presented in each cell of the matrix (Sydoruk, 2019). Furthermore, Webb, 1999 p.3 specified what training reviewers need if they are to validly code assessments. Reviewers were given the following levels to judge depth of knowledge for both mathematics and science (Webb, 1999):

1. Recall: Recall of a fact, information, or procedure.
2. Skill/Concept: Use of information, conceptual knowledge, procedures, two or more steps, etc.
3. Strategic Thinking: Requires reasoning, developing a plan or sequence of steps; has some complexity; more than one possible answer; generally takes less than 10 minutes to do.
4. Extended Thinking: Requires an investigation; time to think and process multiple conditions of the problem or task; and more than 10 minutes to do non-routine manipulations.

In this study, the coding committee discussed each cell of the CRM prior to coding in order to clarify the types of assessment questions that would be placed into each category of the

PARCC assessment. This aided the coding process and afforded the coders the opportunity to align key words and phrases found in the assessment to the CRM. The committee decided that the framework would be more user friendly if the first numbers representing the Webb's Depth of Knowledge levels (columns) were changed to letters, and the second numbers representing Bloom's Taxonomy (rows) would stay the same for each cell. For example [1,2] would now be considered [A,2] in our coding agenda. After the review of the Webb's Alignment Tool, Hess' Cognitive Rigor Matrix and subsequent meetings regarding roles as well as the establishment of the coding agenda, the coding committee took on the task of coding the language arts and mathematics Standards Grades 3 and 4 of the PARCC (2019) practice assessments using the "double-rater read behind consensus model" (Sato et al., 2011 p. 11).

The double-rater method allowed for ongoing consensus during the coding process. For each assessment in each grade level, one analyst independently coded the standards (Sato et al., 2011). A second analyst then reviewed the outcomes of the first analyst's ratings and noted agreement or disagreement with the first analyst's ratings (p.11). Assessment questions were coded in sets of 10 for inter-rater agreement. Questions that the coding committee did not agree upon were marked for later discussion. These discrepancies between the ratings with respect to the criteria and coding dimensions were later discussed and a consensus was reached (Sato et al., 2011). Furthermore, after all the PARCC language arts and mathematics practice assessment were coded, the results of this analysis were compared to similar studies in order to provide consistent methodology in the topic area. Using related studies that coded standards for example (Niebling, 2012; Sato et al., 2011; Sforza, 2014, and Burns, 2017). Furthermore, practices performed in this study modeled those of similar studies in order to provide a consistent methodology on the topic (Burn, 2017; Fitzhugh, 2019; Satos et al., 2011; Sforza, 2014; Sydoruk, 2019).

Data Analysis Procedure

The PARCC practice assessments were analyzed in Grades 3 and 4. In this study I quantified the qualitative data by counting the number of assessment questions that were coded in each cell of the CRM (Creswell, 2009 p. 218). Hess' Cognitive Rigor Matrix was utilized to analyze the cognitive complexity of a standardized assessment. In addition, a cross-reference between Webb's Alignment Tool and the examples given in Hess' Cognitive Rigor Matrix provided the coding committee with the resources to reach an agreement on all questions (Sydoruk, 2019).

The coding committee met on December 16, 2019, in order to discuss and calibrate to the categories found in Hess' Cognitive Rigor Matrix. The two coders initially reviewed both 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. In instances in which coders had an initial disagreement, discrepancies were discussed with respect to the complexity criteria of each question. If a consensus could not be reached, the coders followed the Tips for Facilitating the Consensus Process in the Webb's Alignment Tool manual. The protocol used in this study modeled those of similar studies in order to provide a consistent methodology in the topic area (Miles et al., 2014; Sato, Lagunoff, & Worth, 2011; Sforza, 2014; Sydoruk, 2019).

Following the discussion, coders then used the PARCC Languages Arts Practice Assessment Grade 3 as part of their training and calibration. The two coders completed 55 questions with 100% agreement due to discussion of each question during the training. Utilizing the double-rater read behind method, the coding committee analyzed the test questions. The two coders used a coding table (see Appendix D) to provide a visualization of the categories each test question was placed. It also provided a means of organization, so that the coders can easily check the alignment between them as part of the double-rater method. Figure 8 represents an example of the template use for this study. The completed template can be located in Appendix D.

PARCC Grade 4 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																		

Figure 8. Abridged coding template.

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 59 questions independently. A second coding session was held on January 15, 2020, to review questions in sets of 10 and to discuss any disagreements found with the codes. Furthermore, if there were any disagreements, the coding committee flagged the question so that a discussion could take place and consensus for the categorization on the CRM could be reached. For example, one coder presents the categorization to the other coder using the double-rater read-behind method consensus model. Furthermore, the second coder agrees or challenges using the Hess CRM and the Webb's Alignment Tool to support their choice of categorization. Ultimately, the coding committee followed the suggestion of Webb's et al (. 2005), assigning a higher depth of knowledge level in cases were the coders were not in agreement and were split between two ratings. Questions placed in cells [C,2] to [D,6] are considered higher order thinking.

Data collected were assessed according to frequency and distribution. The total numbers of questions were evaluated in order to calculate percentage. The coding committee reviewed the PARCC Language Arts Practice Assessment Grade 4, completing 59 questions with 92% exact agreement and 100% consensus by the end of the second session. Out of 6 Sets, Sets 1 and 5 were completed with 100% exact agreement. Coders discussed the commonalties between the skills and the types of questions in these sets. In Set 2 the coders agreed on 90% of the questions moving one question from [A,2] to [B,2]. In Set 3, the coders were also in 90% exact agreement moving one question from [B,3] to [B,2]. This question was agreed upon, but was changed to match similar questions. In Set 4, there was 80% agreement, in which the coders discussed the wording of questions. One question was originally placed in [A,2], but due to the key word summarize it was then moved to cell [B,2]. The second question was moved from [A,2] to [B,2] due to locating a description of the main character between two paragraphs in the story. In Set 6, similarly to the question in Set 4 the question included the key word summarize so it was moved from [A,2] to [B,2].

A third calibration session was conducted on January 27, 2020. The double-rater read-behind consensus model was employed again to ensure the reliability of the questions coded. During this session the coders reviewed the PARCC Math Practice Assessment Grade 3. A group of 53 questions in sets of 10 with 81% exact agreement and 100% consensus at the end of the calibration session. In the first set, the coders had 70% total agreement. One question was moved from [A,2] to [B,4] upon discussion of the comparison of data. A consensus was reached regarding the second questions moving from [A,3] to [B,3], due to the multiple steps of adding and subtracting that must be taken to solve the word problem. The last question in this set was increased in cognitive complexity and moved from [B,2] to [C,2]. The word problem expected students to explain their thinking by identifying the incorrect reasoning and providing a correct approach. The second set also had a 70% agreement, similar to a previous question a question was moved from [A,1] to [B,4] comparing data figures. One question in this set was lowered in cognitive complexity and was moved from [B,3] to [A,3] and the other question was moved from [C,4]to [C,2] due to the explanation and generalization of ideas. In Set 3, there was 80% total agreement. Relatedly to a question mentioned previously, the coders increased complexity upon the discussion of comparison of data and figures moving the question from [A,1] to [B,4]. The final question in this set was moved from [A,2] to [B,3] due to the application of multiple concepts. In the fourth set the coders had 90% agreement. The one question in this set was a similar question that had been discussed so, the coders increased cognitive complexity from [A,1] to [B,4]. The fifth and final set of the calibration session had 13 questions with 92% exact agreement, with one question being moved from [A,1] to [A,2], instead of basic recall the question is considered the evaluation of a basic expression.

A fourth and final calibration session was conducted on February 1, 2020. During this session the coders reviewed the PARCC Math Practice Assessment Grade 4. The remaining questions were calibrated in four sets of 10 questions and one set of 13 questions, in which the coding committee had 83% total agreement and 100% consensus at the end of the calibration session. In Set 1, the coders had 80% exact agreement, in which one question was moved from [A,4] to [B,2] and the second question moved from [B,4] to [B,2], the students must explain how they solved the problem as well as show their work using equations. In Set 2, the two coders also reached 80% agreement. In this set, one question was moved from [A,3] to [B,3] and the second question from [A,4] to [B,3]. Coders came to the consensus that both questions were using calculations to figure out the correct equation. Set 3 had 90% agreement, with the question moved similar to past questions from [A,2] to [B,3]. The fourth set has an 80% exact agreement. The coders came to a consensus and agreed that the first question in this set was a higher-order thinking question. This question is an open-ended question that asks students to explain their thinking and to come up with a solution. The second question was lower in cognitive complexity and was moved from [C,2] to [B,3] because it's a routine problem. The fifth and final set had an 85% exact agreement. The first question was lowered from [C,2] to [B,3] and it is similar to the question mentioned in the fourth set. The final question in this set is moved from [A,1] to [B,2], it asks for you to specify which answer is true. The coders came to the consensus that this is not basic recall question and one must solve each question to determine which one is in fact true.

After all data were coded, the frequency and distribution is evaluated in order to calculate a percentage. Similar to Burn's (2017) a formula was then used to calculate the percentages of the cells (DOK/ Blooms Taxonomy level).

$$\frac{\text{\# of PARCC practice test questions coded into cells using the CRM (DOK/Blooms Taxonomy levels)}}{\text{Total \# of practice questions}}$$

For example, if there are 50 questions in the Grade 3 assessment, 8 of which are coded at a level [2,2]. For example, the formula provided the following result:

$$\frac{8}{50} = 16\% \text{ place in cell [2,2] for Grade 3}$$

This data plan analysis addresses the research aforementioned. CRM was used to analyze complex thinking skills, following the use of a basic formula to calculate percentages of level distribution in both the language Arts and mathematics practice test.

Chapter Summary and Subsequent Chapter

Chapter III described the coding protocol used to align 213 English language arts and mathematics questions (including the divided parts of some questions) from the PARCC practice assessment to Hess' Cognitive Rigor Matrix. For this study, qualitative content analysis methodology was employed to code each of the PARCC assessment questions. Furthermore, a quantitative method was then utilized to describe the differences and similarities of complex thinking that exist in the language of the PARCC practice assessment. Webb's Alignment Tool Training manual was used to train the coding committee throughout the process of coding each assessment question within the guidelines previously mentioned. To ensure reliability and validity coders were trained by an experienced coder in deductive coding through calibration exercises. A coding template was subsequently created indicating the assessment questions and its placement beside Hess' Cognitive Rigor Matrix (provided in Appendix A).

Chapter IV presents the findings of this study focusing on the overarching question and the three subquestions.

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 Language Arts and Mathematics Grades 3 and 4. This study aimed to categorize and analyze the frequencies and percentages of complex thinking found in grades 3 and 4 by categorizing the questions found in the assessments. A sample size of 220 questions was used in this mixed method case study. The coding committee held four coding sessions that took place between January 15, 2020 and February 1, 2020. During these sessions the two coders employed Hess' Cognitive Rigor Matrix as an alignment tool to help categorize and determine the higher order thinking that was evident in the PARCC practices assessments. In the matrix, higher order categories included [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. Through the consensus model, the coding committee held calibration sessions to discuss assessment questions. The double-rater read-behind consensus model is regarded as being an effective method for increasing inter-rater reliability (Miles et al., 2014; Sato et al., 2011). The double-rater method allowed for ongoing consensus during the coding process. For each assessment in each grade level analysts coded the questions, reviewed the outcomes, and noted agreement and disagreements. 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 describe the differences and similarities of higher order thinking that exist in the language of the assessment.

A deductive content analysis was employed in this mix methods study, providing the structure needed to study the cognitive complexity within the PARCC assessments. In a deductive content analysis, the organization phase involves categorization matrix where all content is coded for correspondence (Elo et al., 2014). In this particular case study, the Hess' Cognitive Rigor Matrix took both Webb's DOK and Bloom's Taxonomy and developed a framework that allowed the categorization of higher order thinking. According to Creswell (2009), case studies are a strategy of inquiry. Consequently, the PARCC practice assessments employed a categorization matrix that adequately represented the concepts, and from the viewpoint of validity, the matrix accurately captured what was intended (Schreier, 2012 as cited by Elo et al., 2014). The PARCC practice assessments could then be coded to its corresponded category determining its cognitive complexity. The team coded 114 language arts questions and 106 mathematics questions.

The study was guided by the overarching question: *What are the types of thinking are assessed by the questions on 2019 PARCC practice tests in English language arts and mathematics in grades 3 and 4?* Hess' Cognitive Rigor matrix was utilized to assess the thinking requirements of each question in both language arts and math. There were three subquestions that further broke down the overarching question into qualitative and quantitative findings.

Qualitative Findings

Language Arts

The first subquestion 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 3 and 4 associate with the language that promotes higher-order thinking found in research literature?*

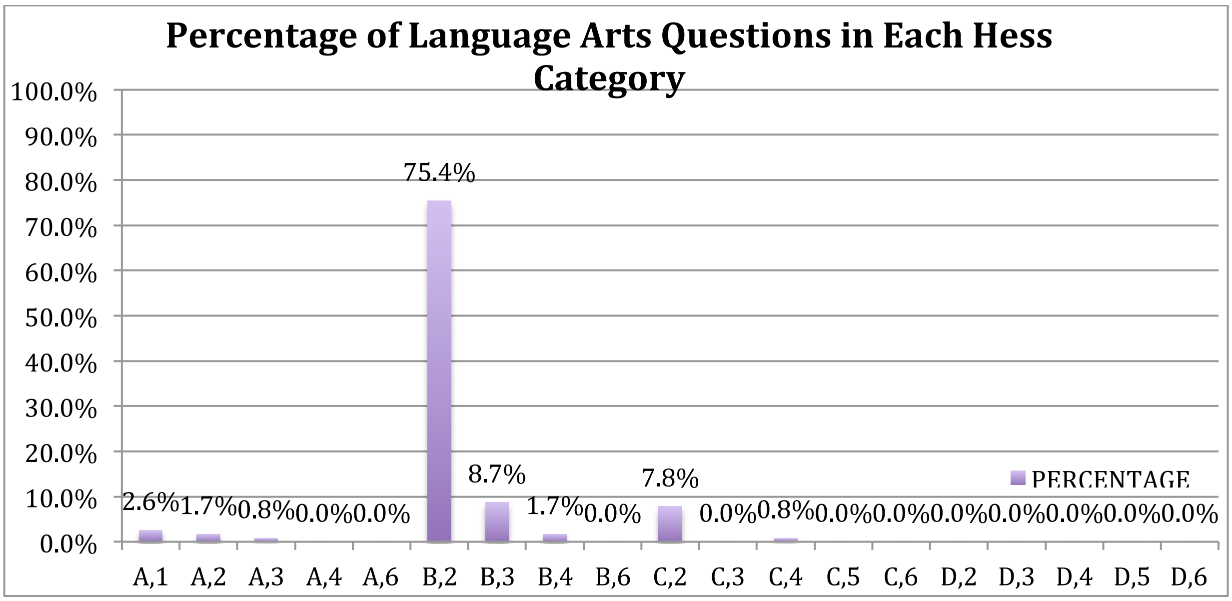

Figure 9. Percentage of language arts questions in each Hess category.

According to Hess' Cognitive Rigor Matrix, the lowest level of cognitive complexity was placed as Level 1 in accordance with Webb's Depth of Knowledge, in which the only expectation is basic recalling, recognizing, and/ or locating basic facts, terms, details events etc. The complexity of the tasks itself increases within Level 1 according to Bloom's Taxonomy levels. For the lowest level of cognitive complexity, Hess' Cognitive Rigor Matrix contains cells [A,1], [A,2], [A,3], [A,4] and [A,6].

Out of 220 questions analyzed, 114 were language arts questions. Three questions in the assessment were placed in [A,1] (Webb's Depth of Knowledge Level 1, Bloom's Taxonomy Recall) cell of the matrix totaling 2.6% 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: *How did Carver become well known across the country?* The answer was directly apparent in the text. Students were asked to select the correct answer consisting facts from the text labeled from A-D.

Two questions were placed into cell [A,2] (Webb's Depth of Knowledge Level 1, Bloom's Taxonomy Understand/ Literal Comprehension), which made up 1.7% of the total number of language arts questions examined. The questions provided reading passages and expected the students to describe facts. For example, one of the provided passages was titled "What Is a Spacewalk?" and the question that followed asked, "What is one kind of important work that astronauts do when they are on a spacewalk". In addition, one question was placed into cell [A,3] (Webb's Depth of Knowledge Level 1, Bloom's Taxonomy Apply), which made up 0.8% of the total number of questions in language arts. This question used word relationships (antonyms), by asking the opposite meaning of *disputing* and providing words to choose from labeled from A-D. The last two cells in Level 1, [A,4] and [A,6] on Hess' Cognitive Rigor Matrix, did not have any questions from the Language Arts PARCC Practice Assessment placed into it.

Level 2 of Hess' Cognitive Rigor Matrix, which is also aligned with Webb's Depth of Knowledge Level 2, "include the engagement of some mental processing beyond a habitual response (Webb, 2007). A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response (p.12). 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 all the questions examined in the language arts assessments, 86 were placed into section [B,2] (Webb's Depth of Knowledge Level 2, Bloom's Taxonomy Understand/ Literal Comprehension). This equated to 75.4% of all questions in the language arts assessments. B,2 was the modal response for the language arts questions. Most of the 86 questions in B,2 required students to identify the main idea and details of the passage, make logical predictions and inferences, summarize ideas, locate information to support central ideas, and/ or explain relationships between the text and the text structure. Many of the questions that fulfilled the criteria for [B,2] asked students questions such as "Which sentence summarizes the speaker's thoughts?", "What is the moral of…", " Which statement provides the best explanation of...", or "Which detail in the passage supports the answer…?" Locating information to support central ideas had the highest frequency of questions placed into cell [B,2], with every question asking students to locate the detail that supports the main idea.

A total of 10 questions, or 8.7% of the language arts questions, were categorized into cell [B,3] (Webb's Depth of Knowledge Level 2, Bloom's Taxonomy Application). The language found in the PARCC Language Arts Grades 3 and 4 encompasses for the questions that aligns to [B,3]. An example of the question which aligns to [B,3] was the following: *What is the meaning of the word drift as it is used in paragraph 18 of "Just Like Home"?* This question asked students to use the context of the text to identify the meaning of a word. Two questions were placed in the [B,4] (Webb's Depth of Knowledge Level 2, Bloom's Taxonomy Analysis) cell of the matrix totaling 1.7% of the Language Arts questions analyzed. An example of a question that includes the language that aligns to [B,4] was the following: *The author of the story "Just Like Home" uses different structural elements than the poet of the poem "Life Doesn't Frighten Me." Which structural element is found only in the story?* This question asked students to compare and analyze the structural elements found in both the poem and story, subsequently finding the element of the story "Just Like Home".

Level 3 requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels (Webb, 2007). Students must explain their thinking 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],and [C,4]. Most of the higher order questions came from the writing tasks in the assessment. The writing questions were open-ended and required explanation and reasoning, using evidence from numerous passages in the assessment.

There were 9 questions placed [C,2] (Webb's Depth of Knowledge Level 3, Bloom's Taxonomy Understand) cell of the CRM, totaling 7.8% of the questions analyzed. An example of a question with language that aligned to [C,2] included the following: *Write and essay to explain what can be learned from the illustrations about the lives of the ponies described in the passages.* This question similar to the other questions placed in this cell asked students to explain, generalize, or connect ideas using supporting evidence from the passages in the assessment. Additionally, one question was placed on [C,4] (Webb's Depth of Knowledge Level 3, Bloom's Taxonomy Analysis) cells, totaling 0.8% of the questions analyzed. Similarly to most of the questions placed in [C,2], the question was open-ended. There were no questions placed in [C,6].

The highest level of cognitive complexity in Hess' Cognitive Rigor Matrix was categorized as Level 4, in correspondence with Webb's Depth of Knowledge Level 4. In this level, requires complex reasoning, planning, developing, and thinking most likely 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].

Mathematics

The second subquestion was: *In what way(s) does the language of the questions on the mathematics section of 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 compare with the language that promotes higher-order thinking found in research literature?*

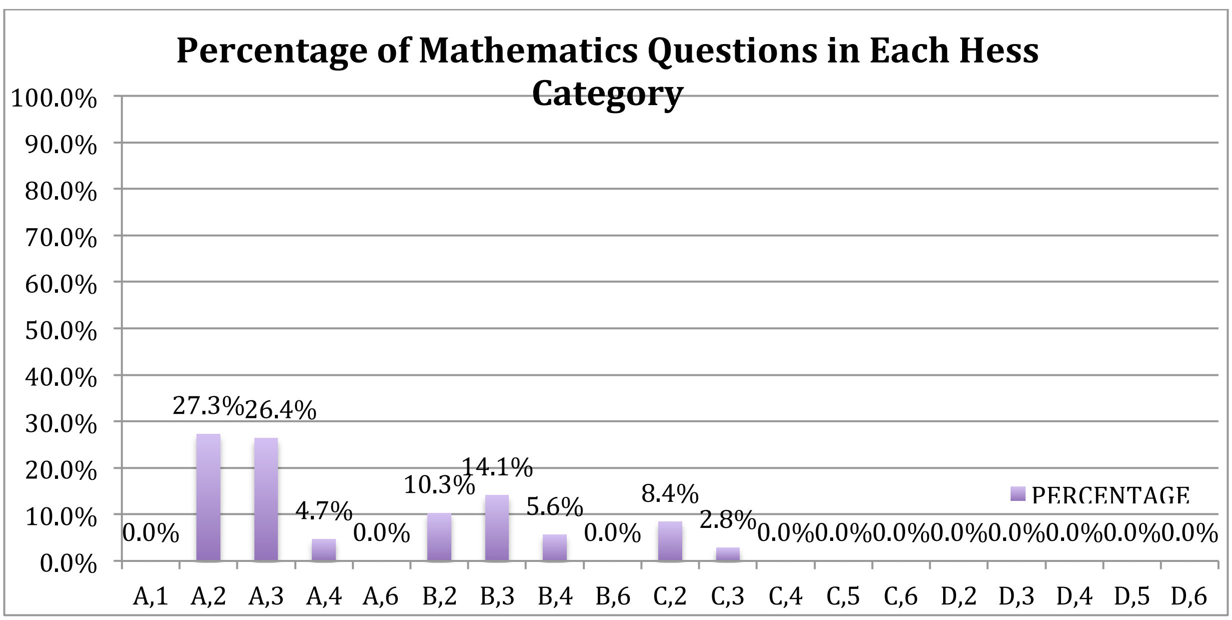

Figure 10. Percentage of mathematics questions in each Hess category.

Hess' Cognitive Rigor Matrix for Math and Science was utilized to assess the thinking requirements in the PARCC Mathematics Practice Test in Grades 3 and 4. As mentioned previously, the lowest level of cognitive complexity contains 5 cells, and amasses for a large number questions in the math practice tests. Out of 106 questions examined in the math tests, 62 equating to 58.4% of all questions were placed into the Level 1 cells. There were no questions placed in [A,1], but 29 questions were placed in [A,2] of the cell matrix totaling 27.3% of the math questions analyzed. Almost all the questions required students to evaluate an expression with the exception of some questions requiring students to locate points on a number line. An example of an [A,2] question found in the math practice test was the following: *Select the three equations that are correct.* The question provided 5 expressions labeled from A-E. An additional question placed in [A,2] asked: *Which number line shows the correct location of the number 5/3?* This question included 4 number lines, labeled A-D with points placed representing a fraction.

Similarly, 28 questions were placed in [A,3] of the CRM totaling 26.4%. The questions placed in the cell mostly expected the students to calculate, and apply algorithm or formula. An example of an [A,3] question found in the test was the following: *What number should replace the ? to make a fraction equivalent to 0.5* The question expects you to calculate and provide a fraction that is equal to 0.5. Another question placed in [A,2] was the following: *Gina's bedroom floor is in the shape of a rectangle. It is 10 feet long and 9 feet wide. What is the area of Gina's bedroom floor?* This question included 4 square footages, labeled A-D. Additionally, five questions were placed into cell [A,4], which made up 4.7% of the total number of questions in the math assessment. The questions required the student to retrieve information from a table or graph to answer the question. The last cell in Level 1, [A,6], did not have any questions.

The questions in Level 2 required students to make some decisions as to how to approach a problem. 11 questions were placed into section [B,2] making up 10.3% of the total questions in the math test . The questions placed on this cell required students to specify and explain relationships, explain steps followed, make basic inferences from data, and use models /diagrams to represent or explain mathematical concepts. An example of a question that included the language that aligns to [B,2] was the following: *Martin gave 1/3 of the corn bread to his neighbor. Explain how you can use the model to show 1/3. Then write a fraction that is equivalent to 1/3.* The student must explain how to use the model to explain mathematical concept. An additional example that also aligns included the following: *Case has 4 boxes. He puts 9 model cars in each box. What is the total number of model cars Cade put in these boxes?* Students must summarize results by solving the mathematical problem.

A total of 15 questions, or 14.1% of the math questions, placed into cell [B,3]. The language found in most of the questions that aligns to [B,3] are questions that solve routine problems and using tables or graphs to retrieve information and use it to solve a problem, both requiring multiple steps. An example of a question that aligns was the following: *Find the total number of pets the fourth grade students have. Explain how you used the bar graph to solve the problem. Show your work using equations.* Six questions were placed in the [B,4] cell of the matrix totaling 5.6% of the math questions analyzed. An example of a question that includes the language that aligns to [B,4] is: *Sandy draws a shape. She divides it into parts. Each part is 1/8 the area of the shape. Which shape could be the one Sandy draws?* The questions placed in this particular cell in both grades 3 and 4 expected students to compare both data and figures. There were no questions placed in cell [B,6].

Level 3, considered higher-level thinking a open-ended and require students to demonstrate their knowledge by explanation, reasoning, using evidence to find mathematical solutions. There were 9 questions placed [C,2] cell of the CRM, totaling 8.4% of the questions analyzed. An example of a question with language that aligned to [C,2] included the following: *Cindy is finding the quotient… She says, "The answer is 18 because addition is the opposite of division and 9+18=27. Identify the incorrect reasoning in Cindy's statement.* This question similar to the other questions placed on this cell asked students to explain, generalize, or connect ideas using supporting evidence to find a solution to the problem. Furthermore, 3 questions was placed on [C,3], totaling 2.8% of the questions analyzed. An example of a question with language that aligned to [C,3] was the following: *Tori uses the expression… Leo uses the expression… Find the total area, in square feet, of the new, larger tabletop. Use the grid to explain why both Tori's expression and Leo's expression are correct.* The question asked for students to analyze data to solve the expressions finding the total area in square feet, then examining the procedures and solutions used to solve. There were no questions placed in cells [C,4], [C,5], and [C,6]. None of the questions examined from the PARCC Math Practice Tests in Grades 3 and 4 placed into Level 4, cells [D,2], [D,3], [D,4], [D,5], and [D,6].

Quantitative Findings

The third subquestion was: *What is the distribution of thinking on the 2016 Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in English language Arts and Mathematics in Grades 3 and 4?*

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 114 were language arts questions and 106 were math questions (see Tables 1 & 2).

Table 1

Distribution of Thinking Requirements on the Language Arts Practice Tests of Grades 3 & 4

DOK 1 Remember	DOK 1 Understand	DOK 1 Apply	DOK 2 Understand	DOK 2 Apply	DOK 2 Analyze	DOK 3 Understand	DOK 3 Analyze
3	2	1	86	10	2	9	1

From the questions analyzed 198 included languages that aligned with lower level thinking (see Figures 9 & 10).

Table 2

Distribution of Thinking Requirements on the Mathematics Tests of a Grades 3 & 4

DOK 1 Understand	DOK 1 Apply	DOK 1 Analyze	DOK 2 Understand	DOK 2 Apply	DOK 2 Analyze	DOK 3 Understand	DOK 3 Apply
29	28	5	11	15	6	9	3

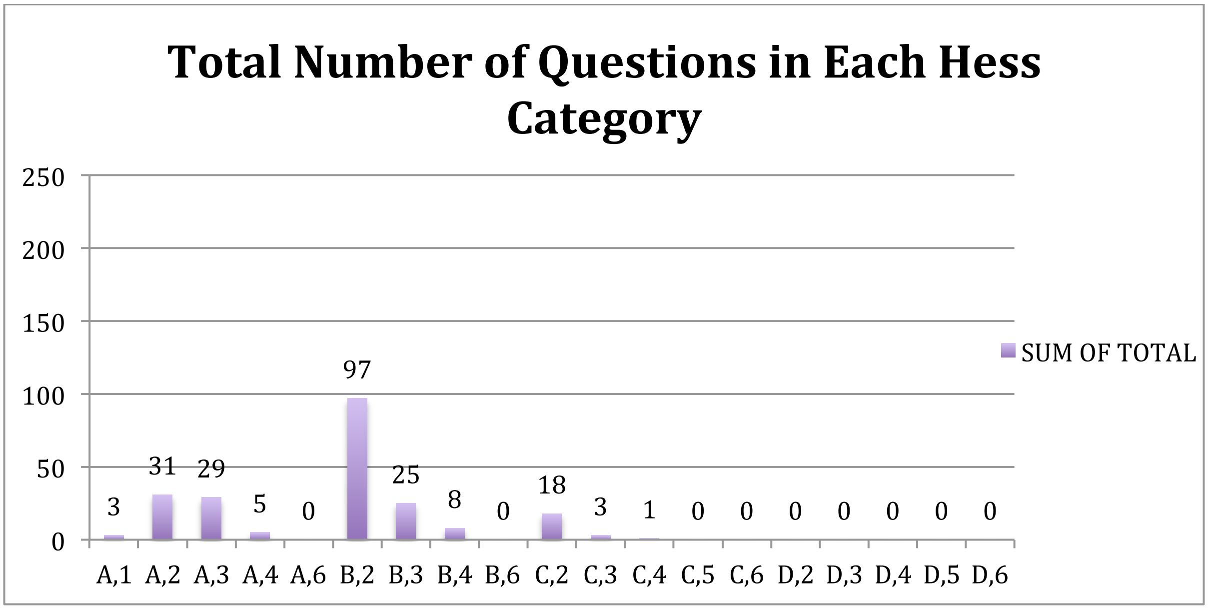

Figure 11. Total number of questions in each Hess category.

The cell with the greatest frequency was [B,2], that had 97 questions making up 44.1% of the total questions analyzed.

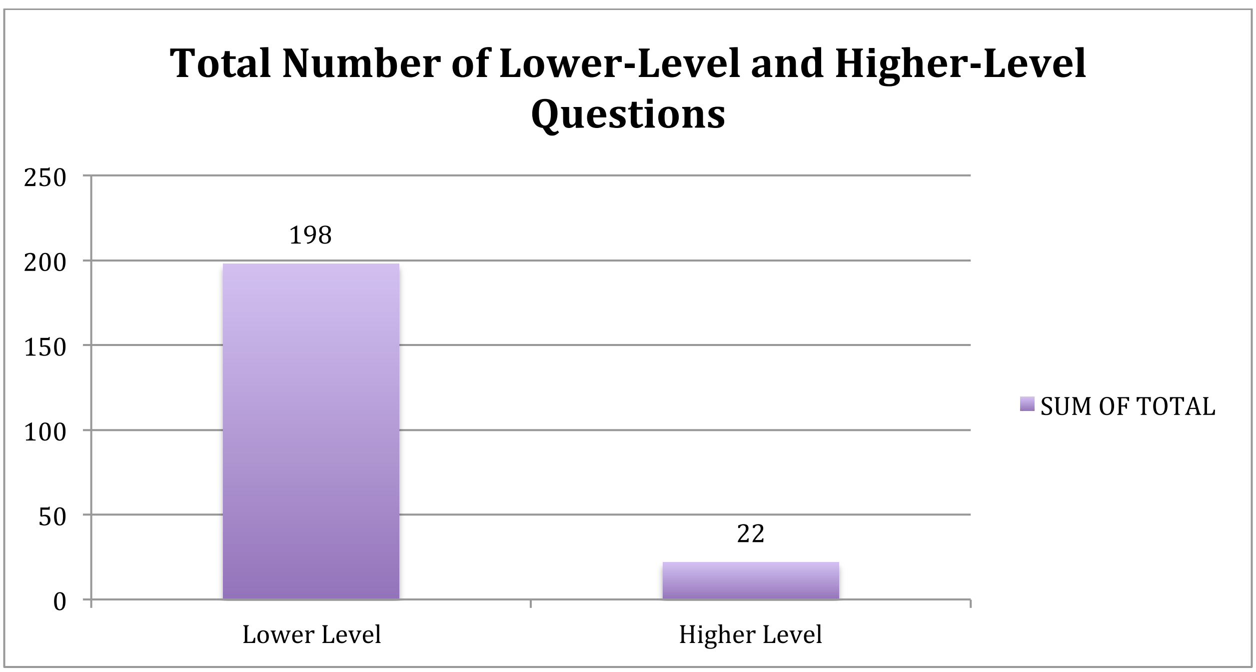

Figure 12. Total number of lower-level and higher-level questions.

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 requires students to demonstrate past a rote response or beyond basic recall like in Level 1

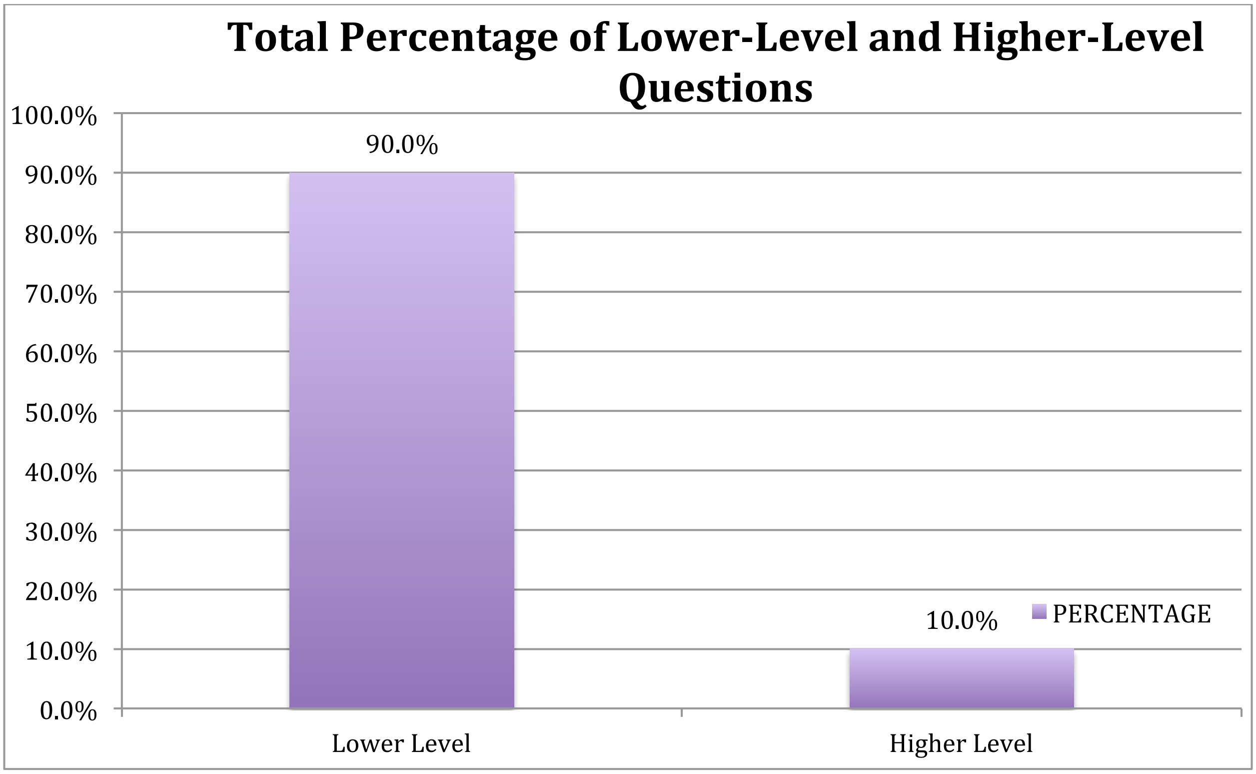

Figure 13: Total Percentage of Lower-Level and Higher-Level Questions

Ninety percent of the questions were categorized as lower level questions requiring students to recall, reproduce, and use skills, and/ or concepts and 10% of the questions analyzed were categorized as cognitive complex requiring strategic thinking, reasoning, and extended thinking.

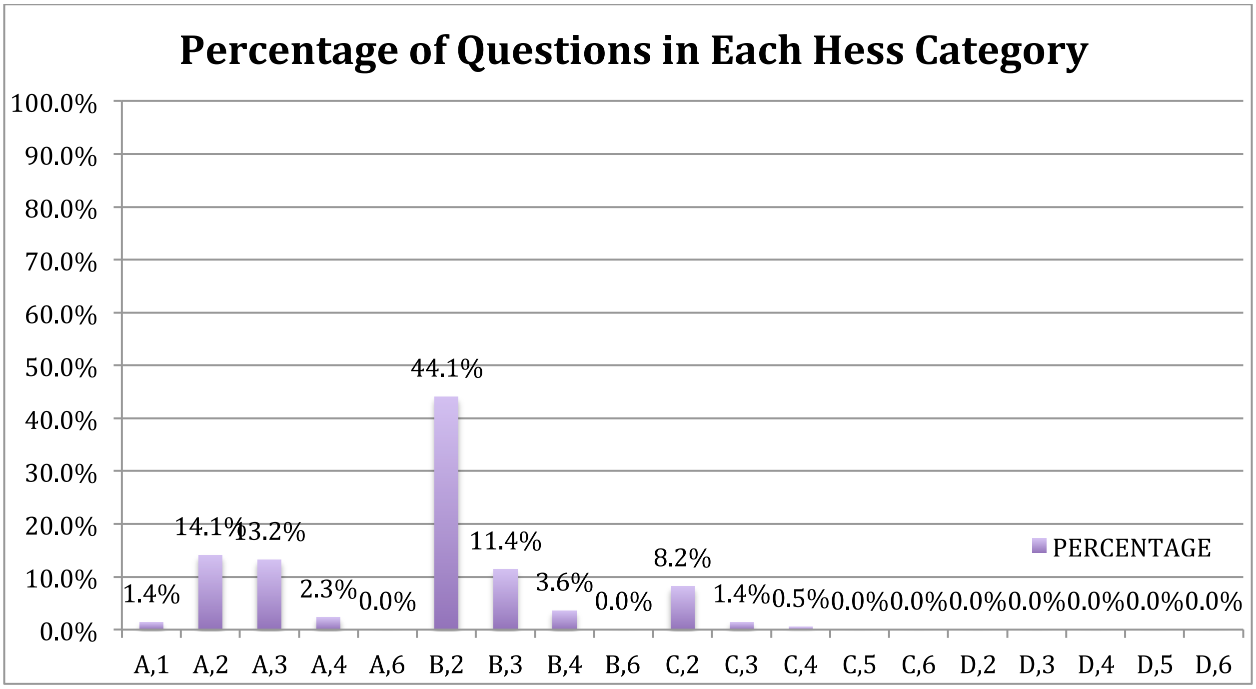

Figure 14: Percentage of Questions in Each Hess Category

Conclusion

The purpose for this mixed methods study was to describe the way(s) in which 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 Grades 3 and 4 associate with the language that promotes higher-order thinking found in research literature. The Elementary grades were selected for this study due to the lack of existing research. Questions that were placed in Levels 3 and 4 of Hess' Cognitive Rigor Matrix were determined to be cognitively complex questions that required students to engage in skills that required higher-order thinking.

In response to the research questions, the data analyzed revealed the following trends from the 220 questions taken from the Language Arts and Mathematics PARCC Practice Tests in grades 3 and 4:

- Out of 220 questions, 198 were categorized as lower questions equating to 90%, of all questions analyzed.
- Out of 220 questions, 22 were categorized as higher-level questions equating to 10% of all questions analyzed.
- The cell with the highest level of frequency was [B,2] which had 97 questions making up 76.3% 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 subquestions, as well as implications for policy and practice, and future research recommendations.

Chapter V: Conclusion

This Mixed-methods study aimed to categorize and analyze the frequencies and percentages of complex thinking in the PARCC practices assessments grades 3 and 4. The purpose for this Mixed-methods study was to describe the way(s) in which 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 Grades 3 and 4 associate with the language that promotes higher-order thinking found in the research literature. The Elementary grades were selected for this study due to the lack of existing research. A sample of 220 English language Arts and mathematics questions was examined using Hess' Cognitive Rigor Matrix. The theoretical framework employed in this study was Hess' Cognitive Rigor Matrix, which superimposes Webb's Depth of Knowledge and Bloom's Taxonomy. Bloom's Taxonomy and Webb's DOK differ in scope and application. According to Hess et al. (2009a), depth of knowledge relates more closely to the depth of content understanding and the scope of learning. Bloom's Taxonomy categorizes the cognitive skills required of the brain to perform a task, describing the type of thinking necessary to complete tasks. Webb's Depth of Knowledge contains four levels, which categorizes the cognitive complexities of tasks. At the lowest level (Level 1), there is a lack of cognitive complexity. The expectation in Level 1 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 (Webb, 2002). The level 2 (Skills/Concepts) is considered lower level in the framework and includes the engagement of some mental processing beyond a habitual response (p5). Levels 3 and 4 are both considered higher-order because they both require cognitively complex tasks that must extend the students thinking and understanding.

Bloom's Taxonomy and Webb's DOK differ in terms of application. According to Hess, et al. (2009b), Bloom's Taxonomy categorizes the cognitive skills required of the brain to perform a task. Subsequently, Bloom's Taxonomy and DOK differ in scope, so Hess' Cognitive Rigor Matrix superimposes both cognitive complexity measures – Bloom's Taxonomy and Webb's Depth of Knowledge – to produce a means of analyzing the emphasis placed on curricular materials, instructional focus, and classroom assessment (Hess, 2013). The levels of Bloom's Taxonomy embedded into the CRM include, from lowest to highest level: remember, understand, apply, analyze, evaluate, and create. The levels can increase in conjunction with the DOK levels or could place at a higher level compared to a lower level in Webb's DOK. These levels are all based on the types of tasks or questions that the students are asked to complete.

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 3 and 4 based on pre-existing codes. 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 Grades 3 and 4 PARCC practice assessments in English Language Arts and Mathematics based on the Hess Cognitive Rigor Matrix methodology. Deductive category application was utilized to connect the language from Hess' Cognitive Rigor Matrix to the language of the 220 English Language Arts and Mathematics questions. The committee decided that the framework was more user friendly if the first numbers representing Webb's Depth of Knowledge levels (columns) were changed to letters, and the second numbers representing Bloom's Taxonomy (rows) would stay the same for each cell. For example [1,2] was considered [A,2] in our coding agenda. To ensure reliability and validity the double-rater read behind method was utilized, the coding committee held sessions in order to calibrate and analyze the assessment questions in sets of 10 questions.

Quantitative methods were utilized in the second part of the study to describe the differences and similarities of higher order thinking that exist in the language of the PARCC practice assessment. The quantitative data collection and analysis built the results of the qualitative phase. All data in the study was coded, the frequency and distribution was then evaluated in order to calculate percentages.

Conclusion

The Partnership for Assessment of Readiness for College and Careers claimed that the PARCC tests assessed students' performance while providing information regarding what students need to learn to be ready for the next grade level, college, and career. Upon analyzing the questions from the test, the PARCC consortia claims do not hold true, as the findings suggest that the assessment questions are not cognitively complex, providing mainly lower level questions. The findings provided a glimpse on what students are assessed. Educational stakeholders have been caught up on the benefits these assessments can provide but have failed to question and further investigate whether or not the claims of higher order thinking in assessments are in fact true. It is important that the standardized assessments are inspected at a state and local level so that the quality of the questions being presented are assessing the students according to instruction with the right amount of cognitively complex questions embedded. School officials need to be more observant, reflective and critical when developing assessments before implementing.

Assessments serve a variety of important functions, including as barometer, resource, and engine of learning (Darling-Hammond, 2015). In order to serve these functions effectively, policymakers must ensure every assessment serves a clear purpose and fits within a broader state and local assessment strategy designed to effectively support learning (Darling-Hammond, 2015). The pressure of performing well in these assessments causes a domino effect. School districts tend to react in a way that in turn places major pressure on educators. Ultimately affecting the quality of instruction. Despite such shortfalls, state policymakers should prioritize investments in higher-quality assessments for one simple reason: States currently spend only about a quarter of 1% of total K-12 education expenditures (about $25 per pupil for NCLB required reading and math tests) on standardized assessments. Many of those assessments have been found to lack higher order thinking and to be of relatively low quality, yet policy makers and some school personnel base many decisions on these tests, thus focusing schools' efforts almost exclusively on low-level skills (Darling-Hammond, 2015).

Lewis and Smith (1993) stated, higher-order thinking occurs when a person takes new information and information stored in memory and relates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations. Higher order thinking questions encourage students to think beyond literal questions. These questions promote critical thinking skills expecting students to apply, analyze, synthesize, and evaluate information instead of simply recalling facts.

The findings show that 90% of the questions in the assessments were considered lower level. Duncker (1945) explained functional fixedness, where subjects are hindered in reaching the solution to a problem by their knowledge of an object's conventional function. The PARCC assessments promote functional fixedness, keeping students in a fixed mindset instead of extending their abilities and into having a growth mindset. The PARCC questions do not allow students to move to an extended way of thinking as evidenced in DOK levels 3 and 4.

Functional fixedness consists in focusing on some function of an object while overlooking another necessary function for problem solving. As aforementioned, PARCC assessments have had negative effects in how teachers react to these assessments. In the classroom the teacher tends to expose students to similar questions found on the test. The pressure of observations, SGO's (student growth objective) and SGP's (student growth percentile) dictates their daily instruction.

The findings of this study revealed that the PARCC assessments must be revised to better serve the students in connection to higher-order learning and assessing. Ultimately, if the operational PARCC questions mirror the practice tests, as PARCC claims then school districts, through policy implications must ensure that the proper groundwork is set. This allows educational stakeholders to properly develop assessments including research-based questions that properly assess students in all facets. This may include both lower-level and higher-order questions providing a glimpse of the student's capabilities at their specific grade level. Furthermore, school leaders will be tasked the job of ensuring that instructional opportunities are put into place that expands the students' creativity and including critical thinking through multiple measures.

Recommendations for Leadership Practice

The findings of this study suggest the importance of knowledge regarding cognitive complexity as it is applied in education. The application of higher-order thinking raises awareness as it pertains to schools. The understanding of Hess' Cognitive Rigor matrix or other frameworks alike, in turn promotes higher order thinking in the classrooms. The onus is on school personnel to evaluate such frameworks. These theoretical frameworks can further the development of curriculum and assessments, which can then properly evaluate students.

Furthermore, the responsibility rests on administrators to provide and engage in professional development that can further the knowledge of higher order implementation. Consequently, educators would utilize frameworks to develop curriculum and assessments.

Zhao (2012) stated that in the pursuit of efficiency, equity, and national consistency, learning standards and curricula essentially homogenize children's learning, serving the same educational diet within a nation (p. 11.) Educational stakeholders must take on the responsibility of ensuring that policy, assessments, curricula, and programs include complex thinking skills. Making policy makers, researcher, education leaders, and assessment developers aware that what matters in education assessment is a wicked problem that cannot be easily solved following traditional approaches (Emler, Zhao, Deng, Yin, & Wang, 2019). A barrier to success, particularly for schools serving marginalized students, is the omnipresence of high stakes, mandated tests across many industrialized nations, which act as a disincentive to higher level thinking (Nehring et al., 2019). As the policy world increasingly relies on tests as a public accountability metric, there is evidence that schools under pressure for test performance narrow curriculum and instruction in order to boost scores (Nehring et al., p.9). Furthermore, educators tend to look at assessments more then the actual standards and the assessments consequently, letting these scores drive instruction in the classroom. Assessments that include a preponderance of lower level thinking influence lower level thinking in the classroom.

Educational stakeholders must take a collaborative approach to solve the lack of higher order thinking implemented in the school level. Educators at the local level have it within their power to defy standardization and change the trajectory of education for millions of students (Tienken, 2016). The key to supporting a collaborative approach is to establish purposeful support structures and agreed-upon processes for encouraging open problem solving (Zhao, Wehmeyer, Basham, & Hansen, 2019). The decision about what is implemented or what is assessed should be determined and discussed by those affected. Stakeholders of education outcomes include students, parents, teachers, school leaders, employers, the public, and policy makers (Zhao et al., 2019).

Within the profession (e.g., teacher training, professional development, school and system leadership), communities of practice are well suited to the fostering of complex skills because they bring practitioners into contact with one another through the medium of inquiry and reflective dialogue (McLaughlin & Talbert, 2006; Wenger, McDermott, & Snyder, 2002 as cited by Nehring et al., 2019). In addition, instructional demand and assessment must work together, if assessments are simplistic, instruction will tend to follow (Nehring et al., p.24). Furthermore, school boards can strengthen their roles by reviewing policies, clarifying goals, and practices, implementing procedures, undertaking more systematic training, and partnering with teacher and administrator organizations to influence state education policies, rather than react to state-generated proposals (Hadderman, 1988 as cited by Burns, 2017). Finally, local school boards must financially support established curricula, supplemental programs, and related teacher trainings required for developing students as complex thinkers (Burns, 2017).

Schools are facing increasing pressure to produce good employees and thus are working hard at what is believed to produce good employees with prescribed standardized curricula, lock-step pacing guides, and standardized tests (Zhao, 2012). Preparing students for life in the 21st century is a complicated endeavor. Globalization, technology, migration, international competition, changing markets, and transnational environmental and political challenges add a new urgency to develop the skills and knowledge students' need for success in the 21st century context (Saavedra & Opfer, 2012). School administrators must apply higher order thinking to all facets of curricula and assessment while also considering the importance of lower order skills. Lower order exercises already exist and are common in existing curricula and should be developed simultaneously with higher order thinking skills (Saavedra & Opfer, 2012). Administrators should work with teachers to first determine whether the PARCC test is narrowing the thinking taking place in classrooms and then facilitate a balance of thinking in the classrooms in schools in which a preponderance of lower-level thinking is taking place.

Administrators can facilitate the implementation of critical thinking into the curriculum. Paul and Nosich (1992) were commissioned by the U.S. Department of Education and the National Center of Educational Statistics provide guidance on how to infuse higher order thinking into assessments at the national level. Their report (a) identified 21 criteria for higher order skills testing, (b) developed a concept of critical thinking that meets these criteria, (c) identified 4 domains of critical thinking, and (d) recommended ways to measure the 4 kinds of critical thinking skills (Paul and Nosich, 1992). The criteria should be considered when implement higher order thinking in schools.

The authors created 21 questions from the criteria to guide assessment development:

1. Can it be used for information processing skills?
2. Can it be used to test flexible skills and abilities that can be used in a wide variety of subjects, situations, contexts, and educational levels?
3. Can it account for important differences among the subject areas?
4. Can it be used to focus on fundamental abilities fitted to the accelerating pace of change and embedded in intellectual history?
5. Can it be used to improve instruction?

6. Can it make clear the interconnectedness of our knowledge and abilities, and why expertise in one area cannot be divorced either from findings in other area or from a sensitivity to the need for interdisciplinary integration?
7. Can it be used to assess those versatile and fundamental skills essential to being a responsible, decision-making member of the workplace?
8. Can it generate clear concepts and well thought-out, rationally articulated goals, criteria and standards?
9. Can it account for the integration of adult-level communication skills, problem- solving, and critical thinking, and legitimately assess all of them without compromising essential features of any of them?
10. Does it respect cultural diversity by focusing on the common-core skills, abilities and traits useful in all cultures?
11. Does it test for thinking that promotes (to quote the September 1991 Kappan) "the active engagement of students in constructing their own knowledge and understanding?"
12. Does it concentrate on assessing the fundamental cognitive structures of communication?
13. Can it be used to assess the central features of making rational decisions as a citizen, a consumer, and a part of a world economy?
14. Can it avoid reducing a complex whole to oversimplified parts?
15. Can it articulate what is central to basic skills for the future?
16. Can it provide the kind of skills that are seen as valuable outside the school as well as inside it?

17 and 18. Can critical thinking be assessed in a way that requires evaluation of authentic problems in realistic contexts where the abilities assessed include those of formulating the problem and initial screening of plausible solutions?

19. Can critical thinking be assessed nationally in a way that is financially affordable?

20 and 21. Can critical thinking be assessed so as to gauge the improvement of students over the course of their education and to measure the achievement of students against national standards? (pp. 5–8)

Aikin's (1942) publication of the landmark *Eight-Year Study* emphasized five critical principles essential in the development of complex thinking: (1) strong emphasis on the student, (2) personal experiences, (3) different development styles, (4) problem solving and making prior knowledge connections, and (5) the ability to approach problems through many different lenses. Creativity must also be considered a facet of higher order thinking. Although some references do not explicitly include creativity as higher order thinking, the very act of generating solutions to problems requires the creative process of going beyond previously learned concepts and rules (King, Goodson, & Rohani, 1998).

King, Goodson, and Rohani (1998) outlines the major features of creativity:

- Creativity involves the consistent use of basic principles or rules in new situations, such as Benjamin Franklin's application of conservation and equilibrium (Crowl et al., 1997).
- Creativity involves discovering and solving problems. Innovative approaches are used to accurately evaluate shortcomings, and actions are taken to remedy those weaknesses (Crowl et al., 1997).
- Creativity involves selecting the relevant aspects of a problem and putting pieces together into a coherent system that integrates the new information with what a person already

knows (Sternberg & Davidson, 1995; Crowl et al., 1997). In a basic sense, it involves a series of decision-making choices between “two or more competing alternatives of action,” each having “several pros and cons associated with it” (Crowl et al., 1997, p. 169).

- Creativity overlaps with other characteristics, such as “intelligence, academic ability, dependability, adaptiveness, and independence” and can “evolve within each of the seven intelligences” (Crowl et al., 1997, pp. 195–196).
- Creativity requires many of the same conditions for learning as other higher order thinking skills. The learning processes are enhanced by supportive environments and deteriorate with fears, insecurities, and low self-esteem. Creativity deteriorates with extrinsic motivation, restraint on choice, and the pressure of outside evaluation (Crowl et al., 1997). (King, Goodson, & Rohani, 1998)

The role of the administrator is to help educators develop their own commitment of promoting complex thinking. To promote higher order thinking, administrators should facilitate professional development for teachers, and support them by providing materials that will help with the transition and implementation of instructional strategies. Principals should establish standards for implementation. The following are suggested recommendations that administrators should facilitate for faculty members:

- Make only minimal use of the lecture method of teaching. Instead, plan for student presentations, debates, and discussions.
- Teaching students to think should be a part of regular class subjects and not a class by itself. Students should be learning and using thinking skills throughout the school day .

- Require interaction between teacher and student, or cooperative learning among students. Student participation, teacher encouragement, and peer interaction correlate positively with improved critical thinking scores.
- Teachers should present both their own opinions and those of others, and should distinguish clearly between the two.
- Teachers should modify tradition al grading practices in the interest of creating a non-judgmental classroom atmosphere of mutual trust and support wherein students feel free to truly express themselves. It is necessary to establish a favorable tone in the classroom that encourages students to open up and become involved, and that does not punish risk-taking behavior.
- Teachers should streamline their course material to allow time for the interactive methods that stimulate students to think.
- Students must also learn to express themselves clearly in writing. Teachers should design communication assignments that encourage thinking-examples, summaries, problem-solving exercises, journal entries (Young, 2012).

In order to provide effective educational preparation for students built upon experiential learning, school administrators need to begin with the review and revision of current curricula to ensure the inclusion of vital complex thinking skills (Burns, 2017). Administrators and educators should be trained to utilize frameworks such as Hess' Cognitive Rigor Matrix. The proper use of higher order thinking frameworks can prove to be beneficial when developing curriculum and assessments. An example of a professional learning activity would engage educators in identifying levels of rigor as defined in the Depth of Knowledge framework, evaluate and revise the rigor of questions using CRM, and assess and adjust the rigor of activities and assignments using the framework.

Recommendations for State Education Policy

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). Educator's need more than incentive instead they need understanding of what they are trying implement and teach. Educating youth who will become innovators and experts in their fields is a primary policy and educational goal in the 21st century (e.g., Obama, Strategy for American Innovation, 2009 as cited by Richland & Begolli, 2016). Local control should be returned to school districts in order to provide students with a democratic education free from one-size-fits-all learning standards and learn from the Cardinal Principles of Secondary Education (Commission on the Reorganization of Secondary Education, 1918 as cited by Fitzhugh, 2019) Policy makers must support the development of localized curricula. Educators must be able to make the curricula decisions according to student need and their own high expectations. As aforementioned, Aikin's (1942) *Eight-Year Study* already demonstrated that curriculum can be an entirely locally developed project and still produce better results than traditional standardized curricular programs (Tienken, 2011).

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, was enacted in 2002. NCLB represented a significant step forward for our nation's children in many respects, particularly on where students were making progress and where they needed additional support, regardless of race, income, zip code, disability, home language, or background (ESSA, U.S. Department of Education, 2020). The law was scheduled for revision in 2007, and over time, NCLB's prescriptive requirements became increasingly unworkable for schools and educators. Recognizing this fact, in 2010, the Obama administration joined a call from educators and families to create a better law that focused on the clear goal of fully preparing all students for success in college and careers (ESSA, U.S. Department of Education, 2020). ESSA includes provisions that will help to ensure success for students and schools. The law helps to support and grow local innovations, including evidence-based and place-based interventions developed by local leaders and educators (ESSA, U.S. Department of Education, 2020).

ESSA provides flexibility and opportunity for state and local officials to develop curricula that instills complex thinking. This in turn prepares students for the required college and career readiness. With this flexibility comes the responsibility on states to make the proper changes that will cultivate complex thinking in students. Through The Every Student Succeeds Act, educational stakeholders have been offered the much-needed flexibility to pursue the endeavors that are considered vital in the education of students.

Although ESSA provides some flexibility, it fails to include that same flexibility for standardized testing requirements. Most states are required to test students in language arts and mathematics in Grades 3 through 8 and once in high school. Consequently, because of these inflexibilities, teaching to the test has sadly become a norm in which discourages purposeful instruction and stifle children's creativity. Standardized testing forces an emphasis on rote learning instead of critical, creative thinking and diminishes students' natural curiosity and joy for learning (Zhao, 2012, p. 18). If standardized testing is a must, policy investment in high-quality testing (either everyday or larger scale standardized tests) can provide an opportunity for supporting the development of complex thinking skills despite the often cited discrepancy between testing and teaching higher order thinking skills (Richland & Begolli, p.11). Schools need a comprehensive set of broad child-centered policies based on evidence, which embrace differentiation of implementation and foster cognitive diversity (Tienken, 2012).

A commercial or a “canned” assessment program cannot replace the teacher (Tienken, 2019). Teacher assessments result in less time spent on “test prep” and more time spent on learning (Tienken, p.59). Finland, ranked as one of the top developed countries for education, does not use standardized tests to drive academic performance in their schools, as educational policy makers believe these assessments narrow the curriculum and lead to harmful competition (Sahlberg, 2011 as cited by Polleck, & Jeffrey, 2017). The ultimate assessment system already exists in public school classrooms: the teacher. State and federal education leaders should invest in developing teachers’ assessment skills instead of spending millions of dollars outsourcing assessments to commercial companies whose tests do not provide usable information for teaching and learning (Tienken, 2019, p.59).

Lastly, state and local officials must get with the times and keep pace with new ideas about what and how teachers learn. Explicit investments in teacher professional development or research must provide teachers with better strategies for understanding and addressing these cognitively complex, higher order thinking skills within the constraints of students’ limited capacity cognitive processing systems (Richland & Begolli, 2016). State officials should provide the offering of professional development. They should be held accountable to provide research-based professional development on ways to assess and promote higher order thinking. Providing significant amount of professional development will result in a secure teacher that can implement strategies to extend student learning.

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 and Mathematics sections of the Partnership for Assessment of Readiness for College and Careers (PARCC) practice tests in Grades 3 and 4 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.

Fowler (2013) describes the policy process as the sequence of events that occurs when a political system considers different approaches to public problems, adopts one of them, tries it out, and evaluates it (p. 14). It is my recommendation that a committee is established including educational stakeholders. This includes administrators, educators, political officials, and community members (parents). They must be tasked with the job of determining through the review of current standardized assessments how complex thinking is embedded. Consequently, with the proper information assessments can be mended to include the proper amount of higher-order and lower-order thinking questions. As aforementioned, comprehension a part of lower order thinking skills is integral to higher order thinking skills development (King et al., 1998). Higher-order thinking skills are grounded in lower-order skills such as discriminations, simple application and analysis, and cognitive strategies and linked to prior knowledge of subject matter content (vocabulary, procedural knowledge, and reasoning patterns) (King et al., p.35).

In addition to the aforementioned committee, further research comparing the distribution of cognitive complexity within the local curriculums and instruction should be conducted in the language arts and mathematics areas at the elementary grade levels, which have not yet been evaluated. Furthermore, the standards implemented in states must be evaluated and modified to include skills that promote 21st century learning. Moreover, the information gained through analyzing assessments, standards, curriculum, and instruction can expand the research as what was presented in this particular study. The replication of this study or other related studies could also be conducted utilizing conceptual frameworks such as Hess' Cognitive Rigor Matrix to further the development and implementation of higher order thinking in education. As the knowledge required for global citizenship becomes more complex, the systems in which we educate the rising generation must evolve (Nehring et al., 2019). The implementation of old low level instructional practices and standardized assessments will not get us there. Using current research to redefine educational expectations and providing the proper professional development for educators would be a great place to begin.

References

Aikin, W. M. (1942). The story of the eight-year study, with conclusions and recommendations. New York, NY: Harper & Brothers.

Airasian, P. W., & Miranda, H. (2002). The Role of Assessment in the Revised Taxonomy. *Theory into Practice*, *41*(4), 249–254. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=eric&AN=EJ667161&site=ehost-live&authtype=sso&custid=s8475574

Anderson, L.W., & Krathwohl, D. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives. New York, NY: Longman. Armstrong,

Anthanassiou, N., McNett, J. M., & Harvey, C. (2003) Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. Journal of Management Education, 27(5), 533–555. doi:10.1177/1052562903252515

Armstrong, P. (2017). Bloom's taxonomy. Retrieved from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy

Ashby, C. (2009). No Child Left Behind Act: Enhancements in the Department of Education's Review Process Could Improve State Academic Assessments. Report to the Chairman, Committee on Health, Education, Labor, and Pensions, U.S. Senate. GAO-09-911.

Assarroudi, A., Nabavi, F. H., Armat, M. R., Ebadi, A., & Vaismoradi, M. (2018). Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing*, *23*(1), 42–55. https://doi.org/10.1177/1744987117741667

Bieri, J., & Blacker, E. (1956). The generality of cognitive complexity in the perception of people and inkblots. The Journal of Abnormal and Social Psychology, 53(1), 112–117. doi:10.1037/h0049167

Blank, R. (2002). Models for alignment analysis and assistance to states. Council of Chief State School Officers Summary Document. Washington, DC: Council of Chief State School Officers.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Retrieved from https://ebookcentral.proquest.com

Brown, N. J. S., Afflerbach, P. P., & Croninger, R. G. (2014). Assessment of Critical-Analytic Thinking. *Educational Psychology Review*, *26*(4), 543–560. https://doi.org/10.1007/s10648-014-9280-4

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.

Camara, W., & Quenemoen, R. (2012). Defining and Measuring College and Career Readiness and Informing the Development of Performance Level Descriptors (PLDs).

Common Core State Standards (2010). Retrieved from http://www.corestandards.org/the-standards

Common Core State Standards Initiative. (2018). What parents should know. Retrieved from http://www.corestandards.org/ELA-Literacy/

Creswell, J. W. (2009). Qualitative inquiry and research design (3rd ed.). Thousand Oaks, CA: Sage. Cross

Crowl, T. K., Kaminsky, S., & Podell, D. M. (1997). Educational psychology: Windows on teaching. Madison, WI: Brown and Benchmark. Dede,

Da'as, R., Schechter, C., & Qadach, M. (2018). Switching Cognitive Gears: School Leaders' Cognitive Complexity. *NASSP Bulletin*, *102*(3), 181–203. https://doi.org/10.1177/0192636518794297

Darling-Hammond, L. (2015). Performance Assessments : How State Policy Can Advance Assessments for 21st Century Learning By Ace Parsi and, (October 2014), 1–18.

Duncker, K. (1945). On problem solving. Psychological Monographs, 58,5 (Whole No. 270).

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open*, *4*(1), 215824401452263. https://doi.org/10.1177/2158244014522633

Emler, T. E., Zhao, Y., Deng, J., Yin, D., & Wang, Y. (2019). Side Effects of Large-Scale Assessments in Education. *ECNU Review of Education*, *2*(3), 279–296.

Eno, L. P. (2018, October 12). *State of New Jersey Department of Education*. Retrieved from:https://bloximages.chicago2.vip.townnews.com/pressofatlanticcity.com/content/tncms /assets/v3/ editorial/3/92/39236271-6166-5dd8-9e84dd584aa2ff2b/5bc8f0d1bafa7.pdf.

Fitzhugh II, G. (2019). A Comparison of Complex Thinking Required by the Elementary New Jersey Student Learning Standards and Past New Jersey Curriculum Standards

Forster, M. (2004). Higher Order Thinking Skills. *ACER*, *11*(1). Retrieved from https://research.acer.edu.au/resdev/vol11/iss11/1

Fowler, F. C. (2013). Policy studies for educational leaders: An introduction. Boston: Pearson.

Freebody, P. R. (2003). *Qualitative research in education : Interaction and practice*. Retrieved from https://ebookcentral.proquest.com

Galdas, P. (2017). Revisiting bias in qualitative research: Reflections on its relationship with funding and impact. *International Journal of Qualitative Methods*. https://doi.org/10.1177/1609406917748992

Gewertz, C. (2018). PARCC Common-Core Test Gets Federal Approval. *Education Week*. Retrieved from www.edweek.org

Gewertz, C. (2019). New Setback for PARCC as Another State Abandons Test; New Mexico and New Jersey reconsider use of exam. Education Week VO - 38, (18). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=edsgov&AN=edsgcl.571382818&site=eds-live&authtype=sso&custid=s8475574

Herman, J., Linn, R., & National Center for Research on Evaluation, S. and S. T. (2013). On the Road to Assessing Deeper Learning: The Status of Smarter Balanced and PARCC Assessment Consortia. CRESST Report 823. In National Center for Research on Evaluation, Standards, and Student Testing (CRESST). National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herman, J., & Linn, R. (2014). New assessments, new rigor. Educational Leadership, 71(6), 34–37.

Hess, K. (2006). "Exploring cognitive demand in instruction and assessment." [online] available: http://www.nciea.org/publications/DOK_ApplyingWebb_KH08.pdf. National

Hess, K., Jones, B., Carlock, D., & Walkup, J. R. (2009a). Cognitive Rigor: Blending the Strengths of Bloom's Taxonomy and Webb's Depth of Knowledge to Enhance Classroom-level Processes. *Knowledge Creation Diffusion Utilization*, 1–8. Retrieved from http://www.standardsco.com/PDF/Cognitive_Rigor_Paper.pdf

Hess, K., Carlock, D., Jones, B., & Walkup, J. R. (2009b). What exactly do "fewer, clearer, and higher standards" really look like in the classroom? Using a cogntive rigor matrix to analyze curriculum, plan lessons, and implement assessments. Retrieved from http://schools.nyc.gov/NR/rdonlyres/D106125F-FFF0-420E-86D9-254761638C6F/0/HessArticle.pdf

Hess, K. (2013). Cognitive Rigor in Today's Classroom. Measured Progress: Common Core, 3787. Retrieved from http://www.asdn.org/wp-content/uploads/CCAP3010_Cognitive_Rigor_in_Todays_Classroom.pdf

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

International Assembly for Collegiate Business Education. (2016). Bloom's taxonomy of educational objectives and writing intended learning outcomes statements. Retrieved from http://iacbe.org/wp-content/uploads/2017/08/Outcomes-Assessment-Plan-Guidelines.pdf

Irvine, J. (2017). A Comparison of Revised Bloom and Marzano's New Taxonomy of Learning. Research In Higher Education Journal, 33

Ives, B., & Obenchain, K. (2016). Experiential Education in the Classroom and Academic Outcomes: For Those Who Want It All, *29*(1), 61–77.

King, F. J., Goodson, L., & Rohani, F. (1998). Higher Order Thinking Skills. Publication of the Educational Services Program, Now Known as the Center for Advancement of Learning and Assessment. Obtido de: Www.Cala.Fsu.Edu, 1–176. Retrieved from http://www.cala.fsu.edu/files/higher_order_thinking_skills.pdf

Lewis, A., & Smith, D. (2001). Defining Higher Order Thinking.

Mathis, W. (2016). Alice in PARCCland: Does' validity study' really prove the Common Core test is valid? The Washington Post, 1–8.

Measured Progress. (2016, Aug.16). Measured progress awarded contract for "Next-Generation" MCAS. Author. Retrieved from https://www.measuredprogress.org/measured-progress-awarded-contract-for-next-generation-mcas/

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research : A guide to design and implementation*. Retrieved from https://ebookcentral.proquest.com

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). Qualitative data analysis: A methods sourcebook. Thousand Oaks, CA: Sage. Moon,

National Academy of Sciences. (2007). Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering. Washington, DC: The National Academies Press.

National Academy of Sciences. (2010). Rising above the gathering storm, revisited: Rapidly approaching category 5. Washington, DC: The National Academies Press.

Nehring, J. H., Charner-Laird, M., & Szczesiul, S. A. (2019). Redefining Excellence: Teaching in Transition, From Test Performance to 21st Century Skills. *NASSP Bulletin*, *103*(1), 5–31. https://doi.org/10.1177/0192636519830772

Newmann, F. M. (1988, May). Higher order thinking in the high school curriculum. NASSP Bulletin, M(508), 53-64. Nickerson,

Nichols-barrer, I., Place, K., Dillon, E., & Gill, B. (2015). Predictive Validity of MCAS and PARCC: Comparing 10th Grade MCAS Tests to PARCC Integrated Math II, Algebra II, and 10th Grade English Language Arts Tests, (617). Retrieved from http://www.mathematica-mpr.com/our-publications-and-findings/publications/predictive-validity-of-mcas-and-parcc-comparing-10th-grade-mcas-tests-to-parcc-integrated-math-ii

NYC Department of Education. (2018). Introduction to depth of knowledge. Retrieved from https://www.weteachnyc.org/resources/collection/depth-knowledge/

PARCC. (2019). Retrieved from PARTNERSHIP FOR ASSESSMENT OF READINESS FOR COLLEGE AND CAREERS website: https://parcc-assessment.org

Parsi, A., Darling-Hammond, L. (2015). Performance Assessments: How State Policy Can Advance Assessments for 21st Century Learning. White Paper. National Association of State Boards of Education. National Association of State Boards of Education. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=eric&AN=ED562629&site=ehost-live&custid=s8475574

Paul, R., & Nosich, G. M. (1992). A Model for the National Assessment of Higher Order Thinking.

Peters, S., Taylor, D., & Leith, C. (2016). Study shows the common core PARCC test does not determine college. *Seattle Education*.

Peterson, D. S. (2017). Engaging elementary students in higher order talk and writing about text. Journal of Early Childhood Literacy, 0(0), 1–21. https://doi.org/10.1177/1468798417690918

Plotting future without PARCC. (2018, Mar 08). *Courier - News* Retrieved from https://search-proquest-com.ezproxy.shu.edu/docview/2011531956?accountid=13793

Polleck, J. N., & Jeffery, J. V. (2017). Common Core Standards and their Impact on Standardized Test Design : A New York Case Study. *The High School Journal*, *101*(1), 1–26.

Richland, L. E., & Begolli, K. N. (2016). Analogy and Higher Order Thinking. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 160–168. https://doi.org/10.1177/2372732216629795

Rothman, R., & Marion, S. F. (2016). The next generation of state assessment and accountability, *97*(8), 34–37.

Sato, E., Lagunoff, R., & Worth, P. (2011). SMARTER Balanced Assessment Consortium Common Core State Standards analysis: Eligible content for the summative assessment (Final Report). Retrieved from eric.gov?id=ED536959 Schwartz, Saavedra, A. R., & Opfer, V. D. (2012). Learning 21st century skills requires 21st century teaching. Phi Delta Kappan, 94(2), 8–13. doi:10.1177/003172171209400203

Sawchuk, S. (2018). Illinois Will Step Back from PARCC Test. Is New Jersey Next? Education Week. Retrieved from www.edweek.org

Shaughnessy, M. F. (2012). *Critical Thinking and Higher Order Thinking : A Current Perspective*. Hauppauge, NY: Nova Science Publishers, Inc. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=e000xna&AN=541469&site=eds-live&authtype=sso&custid=s8475574

Simpson, A., Mokalled, S., Ellenburg, L. A., & Che, S. M. (2015). A Tool for Rethinking Teachers' Questioning, *20*(5).

Sforza, D. (2014). The Inclusion of Cognitive Complexity: A Content Analysis of New Jersey 's Current and Past Intended Curriculum.

Sforza, D., Kim, E., & Tienken, C. H. (2016). A Comparison of Higher-Order Thinking Between the Common Core State Standards and the 2009 New Jersey Content Standards in High School. AASA Journal of Scholarship and Practice, 12(4). https://doi.org/10.1002/ejoc.201200111

Tienken, C. H. (2011). For the Record: What Education Policy Could Be. *Kappa Delta Pi Record*, *75*(4), 298–314. https://doi.org/10.1080/00131725.2011.602467

Tienken, C. H. (2012). What education policy could be. Kappa Delta Pi Record, 75, 298–314. Tienken,

Tienken, C. H. (2015, January 24). Parking the rhetoric on PARCC [Web log post]. Retrieved from http://christienken.com/2015/01/24/parking-the-rhetoric-on-parcc/ Tienken,

Tienken, C. H. (2019). Parking PARCC Claims in the Dumpster of Failed Reforms. Kappa Delta Pi Record, 55(2), 57–59. https://doi.org/10.1080/00228958.2019.1580981

U.S. Department of Education. (2020). Every Student Succeeds Act (ESSA). Retrieved from https://www.ed.gov/essa?src=rn VanTassel-Baska,

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6). Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2002). An Analysis of the Alignment Between Mathematics Standards and Assessments for Three States.

Webb, N. L. (2007). Issues Related to Judging the Alignment of Curriculum Standards and Assessments. *Applied Measurement in Education*, *20*(1), 7–25. Retrieved from http://10.0.4.183/s15324818ame2001_2

Webb, N. L. (2008). Alignment analysis of extended reading standards and assessments. (Final report). Madison, WI: Department of Public Instruction.

Young, L. E. (1992). Critical Thinking Skills: Definitions, Implications for Implementation. NASSP Bulletin, 76(548), 47–54. https://doi.org/10.1177/019263659207654806

Yuan, K., & Le, V. (2012a). Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests. (WR-967-WFHF). Santa Monica, CA: RAND.

Yuan, K., & Le, V. (2012b). Summary: Measuring Deeper Learning through Cognitively Demanding Test Items: Results from the Analysis of Six National and International Exams. *Rand*, 1–146. https://doi.org/10.1214/07-EJS057

Yuan, K., & Le, V.-N. (2014). Measuring Deeper Learning through Cognitively Demanding Test Items: Results from the Analysis of Six National and International Exams. *RAND Corporation*.

Zhao, Y. (2012). World-class learners. Thousand Oaks, CA: Corwin.

Zhao, Y., Wehmeyer, M., Basham, J., & Hansen, D. (2019). Tackling the Wicked Problem of Measuring What Matters: Framing the Questions. ECNU Review of Education, 2(3), 262–278. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=eric&AN=EJ1230124&site=ehost-live&custid=s8475574

Appendix A: Hess Cognitive Rigor Matrix (Reading CRM)

TOOL 1

HESS COGNITIVE RIGOR MATRIX (READING CRM):

Applying Webb's Depth-of-Knowledge Levels to Bloom's Cognitive Process Dimensions

Revised Bloom's Taxonomy	Webb's DOK Level 1 Recall & Reproduction	Webb's DOK Level 2 Skills & Concepts	Webb's DOK Level 3 Strategic Thinking/Reasoning	Webb's DOK Level 4 Extended Thinking
Remember Retrieve knowledge from long-term memory, recognize, recall, locate, identify	o Recall, recognize, or locate basic facts, terms, details, events, or ideas explicit in texts o Read words orally in connected text with fluency & accuracy	Use these Hess CRM curricular examples with most close reading or listening assignments or assessments in any content area.		
Understand Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion), predict, compare/contrast, match like ideas, explain, construct models	o Identify or describe literary elements (characters, setting, sequence, etc.) o Select appropriate words when intended meaning/definition is clearly evident o Describe/explain who, what, where, when, or how o Define/describe facts, details, terms, principles o Write simple sentences	o Specify, explain, show relationships; explain why (e.g., cause-effect) o Give non-examples/examples o Summarize results, concepts, ideas o Make basic inferences or logical predictions from data or texts o Identify main ideas or accurate generalizations of texts o Locate information to support explicit-implicit central ideas	o Explain, generalize, or connect ideas using supporting evidence (quote, example, text reference) o Identify/ make inferences about explicit or implicit themes o Describe how word choice, point of view, or bias may affect the readers' interpretation of a text o Write multi-paragraph composition for specific purpose, focus, voice, tone, & audience	o Explain how concepts or ideas specifically relate to other content domains (e.g., social, political, historical) or concepts o Develop generalizations of the results obtained or strategies used and apply them to new problem-based 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	o Use language structure (pre/suffix) or word relationships (synonym/antonym) to determine meaning of words o Apply rules or resources to edit spelling, grammar, punctuation, conventions, word use o Apply basic formats for documenting sources	o Use context to identify the meaning of words/phrases o Obtain and interpret information using text features o Develop a text that may be limited to one paragraph o Apply simple organizational structures (paragraph, sentence types) in writing	o Apply a concept in a new context o Revise final draft for meaning or progression of ideas o Apply internal consistency of text organization and structure to composing a full composition o Apply word choice, point of view, style to impact readers' /viewers' interpretation of a text	o Illustrate how multiple themes (historical, geographic, social, artistic, literary) may be interrelated o Select or devise an approach among many alternatives to research a novel problem
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)	o Identify whether specific information is contained in graphic representations (e.g., map, chart, table, graph, T-chart, diagram) or text features (e.g., headings, subheadings, captions) o Decide which text structure is appropriate to audience and purpose	o Categorize/compare literary elements, terms, facts/details, events o Identify use of literary devices o Analyze format, organization, & internal text structure (signal words, transitions, semantic cues) of different texts o Distinguish: relevant-irrelevant information; fact/opinion o Identify characteristic text features; distinguish between texts, genres	o Analyze information within data sets or texts o Analyze interrelationships among concepts, issues, problems o Analyze or interpret author's craft (literary devices, viewpoint, or potential bias) to create or critique a text o Use reasoning, planning, and evidence to support inferences	o Analyze multiple sources of evidence, or multiple works by the same author, or across genres, time periods, themes o Analyze complex/abstract themes, perspectives, concepts o Gather, analyze, and organize multiple information sources o Analyze discourse styles
Evaluate Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique	"UG" – unsubstantiated generalizations = stating an opinion without providing any support for it!		o Cite evidence and develop a logical argument for conjectures o Describe, compare, and contrast solution methods o Verify reasonableness of results o Justify or critique conclusions drawn	o Evaluate relevancy, accuracy, & completeness of information from multiple sources o Apply understanding in a novel way, provide argument or justification for the application
Create Reorganize elements into new patterns/structures, generate, hypothesize, design, plan, produce	o Brainstorm ideas, concepts, problems, or perspectives related to a topic, principle, or concept	o Generate conjectures or hypotheses based on observations or prior knowledge and experience	o Synthesize information within one source or text o Develop a complex model for a given situation o Develop an alternative solution	o Synthesize information across multiple sources or texts o Articulate a new voice, alternate theme, new knowledge or perspective

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 B: Hess Cognitive Rigor Matrix (Math-Science CRM)

TOOL 2

HESS COGNITIVE RIGOR MATRIX (MATH-SCIENCE CRM):

Applying Webb's Depth-of-Knowledge Levels to Bloom's Cognitive Process Dimensions

Revised Bloom's Taxonomy	Webb's DOK Level 1 Recall & Reproduction	Webb's DOK Level 2 Skills & Concepts	Webb's DOK Level 3 Strategic Thinking/Reasoning	Webb's DOK Level 4 Extended Thinking
Remember Retrieve knowledge from long-term memory, recognize, recall, locate, identify	o Recall, observe, & recognize facts, principles, properties o Recall/ identify conversions among representations or numbers (e.g., customary and metric measures)	Use these Hess CRM curricular examples with most mathematics or science assignments or assessments.		
Understand Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion), predict, compare/contrast, match like ideas, explain, construct models	o Evaluate an expression o Locate points on a grid or number on number line o Solve a one-step problem o Represent math relationships in words, pictures, or symbols o Read, write, compare decimals in scientific notation	o Specify and explain relationships (e.g., non-examples/examples; cause-effect) o Make and record observations o Explain steps followed o Summarize results or concepts o Make basic inferences or logical predictions from data/observations o Use models /diagrams to represent or explain mathematical concepts o Make and explain estimates	o Use concepts to solve non-routine problems o Explain, generalize, or connect ideas using supporting evidence o Make and justify conjectures o Explain thinking/reasoning when more than one solution or approach is possible o Explain phenomena in terms of concepts	o Relate mathematical or scientific concepts to other content areas, other domains, or other concepts o Develop generalizations of the results obtained and the strategies used (from investigation or readings) 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	o Follow simple procedures (recipe-type directions) o Calculate, measure, apply a rule (e.g., rounding) o Apply algorithm or formula (e.g., area, perimeter) o Solve linear equations o Make conversions among representations or numbers, or within and between customary and metric measures	o Select a procedure according to criteria and perform it o Solve routine problem applying multiple concepts or decision points o Retrieve information from a table, graph, or figure and use it solve a problem requiring multiple steps o Translate between tables, graphs, words, and symbolic notations (e.g., graph data from a table) o Construct models given criteria	o Design investigation for a specific purpose or research question o Conduct a designed investigation o Use concepts to solve non-routine problems o Use & show reasoning, planning, and evidence o Translate between problem & symbolic notation when not a direct translation	o Select or devise approach among many alternatives to solve a problem o Conduct a project that specifies a problem, identifies solution paths, solves the problem, and reports results
Analyze Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, find coherence, deconstruct	o Retrieve information from a table or graph to answer a question o Identify whether specific information is contained in graphic representations (e.g., table, graph, T-chart, diagram) o Identify a pattern/trend	o Categorize, classify materials, data, figures based on characteristics o Organize or order data o Compare/ contrast figures or data o Select appropriate graph and organize & display data o Interpret data from a simple graph o Extend a pattern	o Compare information within or across data sets or texts o Analyze and draw conclusions from data, citing evidence o Generalize a pattern o Interpret data from complex graph o Analyze similarities/differences between procedures or solutions	o Analyze multiple sources of evidence o Analyze complex/abstract themes o Gather, analyze, and evaluate information
Evaluate Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique	"UG" – unsubstantiated generalizations = stating an opinion without providing any support for it!		o Cite evidence and develop a logical argument for concepts or solutions o Describe, compare, and contrast solution methods o Verify reasonableness of results	o Gather, analyze, & evaluate information to draw conclusions o Apply understanding in a novel way, provide argument or justification for the application
Create Reorganize elements into new patterns/structures, generate, hypothesize, design, plan, produce	o Brainstorm ideas, concepts, or perspectives related to a topic	o Generate conjectures or hypotheses based on observations or prior knowledge and experience	o Synthesize information within one data set, source, or text o Formulate an original problem given a situation o Develop a scientific/mathematical model for a complex situation	o Synthesize information across multiple sources or texts o Design a mathematical model to inform and solve a practical or abstract situation

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 (Writing/ Speaking CRM)

TOOL 3

HESS COGNITIVE RIGOR MATRIX (WRITING/SPEAKING CRM):
Applying Webb's Depth-of-Knowledge Levels to Bloom's Cognitive Process Dimensions

Revised Bloom's Taxonomy	Webb's DOK Level 1 Recall & Reproduction	Webb's DOK Level 2 Skills & Concepts	Webb's DOK Level 3 Strategic Thinking/Reasoning	Webb's DOK Level 4 Extended Thinking
Remember Retrieve knowledge from long-term memory, recognize, recall, locate, identify	o Complete short answer questions with facts, details, terms, principles, etc. (e.g., label parts of diagram)	Use these Hess CRM curricular examples with most writing and oral communication assignments or assessments in any content area.		
Understand Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion), predict, compare/contrast, match like ideas, explain, construct models	o Describe or define facts, details, terms, principles, etc. o Select appropriate word/phrase to use when intended meaning/definition is clearly evident o Write simple complete sentences o Add an appropriate caption to a photo or illustration o Write "fact statements" on a topic (e.g., spiders build webs)	o Specify, explain, show relationships; explain why, cause-effect o Provide and explain non-examples and examples o Take notes; organize ideas/data (e.g., relevance, trends, perspectives) o Summarize results, key concepts, ideas o Explain central ideas or accurate generalizations of texts or topics o Describe steps in a process (e.g., science procedure, how to and why control variables)	o Write a multi-paragraph composition for specific purpose, focus, voice, tone, & audience o Develop and explain opposing perspectives or connect ideas, principles, or concepts using supporting evidence (quote, example, text reference, etc.) o Develop arguments of fact (e.g., Are these criticisms supported by the historical facts? Is this claim or equation true?)	o Use multiple sources to elaborate on how concepts or ideas specifically draw from other content domains or differing concepts (e.g., research paper, arguments of policy – should this law be passed? What will be the impact of this change?) o Develop generalizations about the results obtained or strategies used and apply them to a new problem or contextual scenario
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	o Apply rules or use resources to edit specific spelling, grammar, punctuation, conventions, or word use o Apply basic formats for documenting sources	o Use context to identify/infer the intended meaning of words/phrases o Obtain, interpret, & explain information using text features (table, diagram, etc.) o Develop a (brief) text that may be limited to one paragraph, précis o Apply basic organizational structures (paragraph, sentence types, topic sentence, introduction, etc.) in writing	o Revise final draft for meaning, progression of ideas, or logic chain o Apply internal consistency of text organization and structure to a full composition or oral communication o Apply a concept in a new context o Apply word choice, point of view, style, rhetorical devices to impact readers' interpretation of a text	o Select or devise an approach among many alternatives to research and present a novel problem or issue o Illustrate how multiple themes (historical, geographic, social) may be interrelated within a text or topic
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)	o Decide which text structure is appropriate to audience and purpose (e.g., compare-contrast, proposition-support) o Determine appropriate, relevant key words for conducting an Internet search or researching a topic	o Compare/contrast perspectives, events, characters, etc. o Analyze/revise format, organization, & internal text structure (signal words, transitions, semantic cues) of different print and non-print texts o Distinguish: relevant-irrelevant information; fact/opinion (e.g., What are the characteristics of a hero's journey?) o Locate evidence that supports a perspective/differing perspectives	o Analyze interrelationships among concepts/ issues/problems in a text o Analyze impact or use of author's craft (literary devices, viewpoint, dialogue) in a single text o Use reasoning and evidence to generate criteria for making and supporting an argument of judgment (Was FDR a great president? Who was the greatest ball player?) o Support conclusions with evidence	o Analyze multiple sources of evidence, or multiple works by the same author, or across genres, or time periods o Analyze complex/abstract themes, perspectives, concepts o Gather, analyze, and organize multiple information sources o Compare and contrast conflicting judgments or policies (e.g., Supreme Court decisions)
Evaluate Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique	"UG" – unsubstantiated generalizations = stating an opinion without providing any support for it!		o Evaluate validity and relevance of evidence used to develop an argument or support a perspective o Describe, compare, and contrast solution methods o Verify or critique the accuracy, logic, and reasonableness of stated conclusions or assumptions	o Evaluate relevancy, accuracy, & completeness of information across multiple sources o Apply understanding in a novel way, provide argument or justification for the application o Critique the historical impact (policy, writings, discoveries, etc.)
Create Reorganize elements into new patterns/structures, generate, hypothesize, design, plan, produce	o Brainstorm facts, ideas, concepts, problems, or perspectives related to a topic, text, idea, issue, or concept	o Generate conjectures, hypotheses , or predictions based on facts, observations, evidence/observations, or prior knowledge and experience o Generate believable "grounds" (reasons) for an opinion-argument	o Develop a complex model for a given situation or problem o Develop an alternative solution or perspective to one proposed (e.g., debate)	o Synthesize information across multiple sources or texts in order to articulate a new voice, alternate theme, new knowledge or nuanced perspective

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: Coding Table

Language Arts Grade 3

Question		[A,1]	[A,2]	[A,3]	[A,4]	[A,6]	[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	1a							X												
2	1b						X													
3	2a						X													
4	2b						X													
5	3a						X													
6	3b						X													
7	4a						X													
8	4b						X													
9	5a						X													
10	5b						X													
11	6a						X													
12	6b						X													
13	8a							X												
14	8b						X													
15	9a										X									
16	9b						X													
17	10a		X																	
18	10b						X													
19	11a	X																		
20	11b						X													
21	12a							X												
22	12b						X													
23	13a						X													
24	13b						X													
25	14a	X																		
26	14b						X													
27	15a						X													
28	15b						X													

Appendix D: Coding Table

Language Arts Grade 3 (continued)

29	16a							X												
30	16b						X													
31	17a										X									
32	17b						X													
33	19a						X													
34	19b						X													
35	20a						X													
36	20b						X													
37	21a						X													
38	21b						X													
39	22a						X													
40	22b						X													
41	24a							X												
42	24b						X													
43	25a						X													
44	25b						X													
45	26a										X									
46	26b						X													
47	27a						X													
48	27b						X													
49	28a		X																	
50	28b						X													
51	29a						X													
52	29b						X													

Appendix D: Coding Table

Language Arts Grade 4

Question		[A,1]	[A,2]	[A,3]	[A,4]	[A,6]	[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	1a							X												
2	1b						X													
3	2a						X													
4	2b						X													
5	3a						X													
6	3b						X													
7	4a						X													
8	4b						X													
9	5a						X													
10	5b						X													
11	6a								X											
12	6b						X													
13	8a							X												
14	8b			X																
15	9a						X													
16	9b						X													
17	10a						X													
18	10b						X													
19	11a						X													
20	11b						X													
21	12a							X												
22	12b						X													
23	13a						X													
24	13b						X													
25	14a						X													

Appendix D: Coding Table

Language Arts Grade 4 (continued)

26	14b						X													
27	15a						X													
28	15b						X													
29	16a						X													
30	16b						X													
31	17a										X									
32	17b						X													
33	18a							X												
34	18b						X													
35	19a						X													
36	19b						X													
37	21a						X													
38	21b						X													
39	22a						X													
40	22b						X													
41	23a								X											
42	23b						X													
43	24a						X													
44	24b						X													
45	26a							X												
46	26b						X													
47	27a						X													
48	27b						X													
49	28a						X													
50	28b						X													
51	29a						X													
52	29b						X													
53	30a						X													
54	30b						X													

Appendix D: Coding Table

Writing Grade 3

Questions		[A,1]	[A,2]	[A,3]	[A,4]	[A,6]	[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	7												X							
2	18										X									
3	23										X									

Writing Grade 4

Questions		[A,1]	[A,2]	[A,3]	[A,4]	[A,6]	[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	7										X									
2	20										X									
3	25										X									

Appendix D: Coding Table

Math Grade 3

Question		[A,1]	[A,2]	[A,3]	[A,4]	[A,6]	[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		X																	
2	2		X																	
3	3								X											
4	4a							X												
5	4b			X																
6	4c				X															
7	5						X													
8	6			X																
9	7		X																	
10	8a										X									
11	8b										X									
12	9		X																	
13	10a				X															
14	10b								X											
15	11								X											
16	12			X																
17	13		X																	
18	14		X																	
19	15										X									
20	16						X													
21	17						X													
22	18a			X																
23	18b			X																
24	19						X													
25	20			X																
26	21				X															
27	22								X											
28	23							X												
29	24		X																	

Appendix D: Coding Table

Math Grade 3 (continued)

30	25a						X													
31	25b								X											
32	26a							X												
33	26b							X												
34	27			X																
35	28										X									
36	29		X																	
37	30		X																	
38	31		X																	
39	32			X																
40	33a			X																
41	33b										X									
42	34		X																	
43	35							X												
44	36a										X									
45	36b											X								
46	37		X																	
47	38		X																	
48	39			X																
49	40a			X																
50	40b		X																	
51	41		X																	
52	42		X																	
53	43		X																	

Appendix D: Coding Table

Math Grade 4

Question		[A,1]	[A,2]	[A,3]	[A,4]	[A,6]	[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			X																
2	2a			X																
3	2b			X																
4	3		X																	
5	4a				X															
6	4b						X													
7	4c						X													
8	5		X																	
9	6		X																	
10	7		X																	
11	8a			X																
12	8b											X								
13	9		X																	
14	10a							X												
15	10b							X												
16	11		X																	
17	12		X																	
18	13			X																
19	14							X												
20	15		X																	
21	16			X																
22	17a		X																	
23	17b			X																
24	18							X												
25	19								X											
26	20a			X																
27	20b			X																
28	21			X																
29	22			X																

Appendix D: Coding Table

Math Grade 4 (continued)

30	23a							X												
31	23b										X									
32	24a							X												
33	24b							X												
34	25		X																	
35	26a						X													
36	26b						X													
37	27				X															
38	28		X																	
39	29a							X												
40	29b			X																
41	30							X												
42	31a											X								
43	31b						X													
44	32						X													
45	33		X																	
46	34a										X									
47	34b										X									
48	35			X																
49	36			X																
50	37a			X																
51	37b			X																
52	38							X												
53	39			X																