Fall 8-17-2020

An Analysis of the Complex Thinking Requirements of the TerraNova and IOWA Practice Tests in English/Language Arts for Grade 8: A Tale of Two Tests

Armand Lamberti
Seton Hall University, armand.lamberti@student.shu.edu

An Analysis of the Complex Thinking Requirements of the TerraNova and IOWA Practice Tests in English/Language Arts for Grade 8: A Tale of Two Tests

Armand Lamberti

Dissertation Committee

Christopher H. Tienken, Ed.D.
Samuel F. Fancera, Ed.D.
Paige D. Sydoruk, 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

Armand Lamberti 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:
Dr. Christopher Tienken _____ 8-17-20
Date

Committee Member:
Dr. Samuel Fancera _____ 8-17-20
Date

Committee Member:
Dr. Paige Sydoruk _____ 8-17-20
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

The importance of higher-order thinking, 21st-Century Skills, and standardized testing are important issues in education. TerraNova and IOWA tests are widely used to assess students' academic competence. The publishers of the TerraNova and IOWA tests claim that their assessment instruments challenge students to employ higher-order thinking skills. This study sought to examine and describe ways in which the language found in the TerraNova and IOWA English/Language Arts practice tests compared with the language that promotes higher-order thinking found in the literature. A convergent, parallel mixed-methods study with qualitative and quantitative content analysis methods was conducted to (a) describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 compares with the language associated with higher-order thinking found in research literature and (b) to describe and compare the complex thinking requirements found on the TerraNova and the IOWA English/Language Arts Practice Tests for grade 8. The qualitative method consisted of a content analysis of the language of the questions on each test, deductive coding, and categorizing the cognitive level of each question based on Webb's Depth of Knowledge. The quantitative aspect of the study consisted of calculating the percentages of questions categorized in each level of Webb's Depth of Knowledge. Each depth of knowledge level represents a level of cognitive complexity: levels 1 and 2 – lower level, levels 3 and 4 – higher level. Each question was rated on a 1–4 DOK level based on Webb's Depth of Knowledge methodology. To assist with reliability in coding, the coders utilized a double-rater read behind consensus model. This study sought to reveal the extent to which complex/higher-order thinking skills are incorporated throughout the TerraNova and IOWA standardized practice tests. The results of this convergent, parallel mixed-methods study found that 98% of the 220

test questions analyzed are promoting lower level thinking: Level 1 questions accounting for 76%, and Level 2 questions accounting for 22%. Of the questions analyzed and coded, only 3 questions were categorized as DOK Level 3, equating to 2% of all questions analyzed and scored in this study. No questions analyzed in this study were categorized as DOK Level 4, equating to 0% of all questions reviewed. The results suggest that the questions promote functional fixedness as opposed to higher-order thinking. Questions categorized as levels 1 and 2 of Webb's DOK are lower level questions with a focus on declarative and procedural knowledge (recall, reconstruction). A consistent focus on lower level declarative and procedural thinking can stunt the complex thinking development of students and lead to functional fixedness. Functional fixedness is known as the phenomenon of one perceiving an entity as having only one function or use. It is a mental set that incites myopic thinking so that an individual is unable see alternatives.

Keywords/terms: cognitive complexity, creative thinking, higher-order thinking, problem solving, standardized assessment, strategic thinking, Webb's Depth of Knowledge (DOK)

Dedication and Acknowledgements

George Eliot wrote, “It is never too late to be what you might have been”. That famous ‘it’s never too late’ quote quite accurately applies to my doctoral quest. With God-given drive and determination, here I am, just turned 65 years old, Armand Lamberti, Ed.D. That noted, I should have accomplished this in my early 40’s, but stepped away far too long. Or was it too long? The Eliot quote has taken on literal meaning to me. Indeed, it is never too late to be what I might have been – as now, I am.

Although attainment of a doctoral degree has come to be an accomplishment to cross of my bucket list, in some sense attaining it at this juncture in life makes the accomplishment all the greater – all the more satisfying. My faith has played a large role in any accomplishments I have been fortunate enough to experience in a long career as an educator. However, along with faith, family and friends have always been there for support and encouragement. I thank my wife, Chantal, and our ever-interesting and wonderful offspring Armando, Paul, and Mary for always being a true testament to the importance of family, and for all the joy and love they bring to my life. Thanks to all my friends for their friendship.

I cannot thank Dr. Christopher H. Tienken enough for his extremely vital role in my earning an Ed.D. Without his mentorship, advisement, scholarship, expertise, and friendship, I may not have been able to make it through to this doctoral accomplishment. In a career that has seen me view very few individuals as heroes, I have added one more – Dr. Tienken. Dissertation committee members Dr. Paige Sydoruk, and Dr. Samuel Fancera have also played vital roles in the completion of my dissertation. Their expertise and guidance throughout the process are truly appreciated and valued. I am hopeful that I will be in touch with, and perhaps work in some way

with Drs. Tienken, Sydoruk, and Fancera going forward. I thank Dr. Christie Vanderhook for her outstanding work as the expert coder in this study, and for her encouragement and friendship. I wish to also express thanks to all the professors I have had at SHU. Particularly, I want to thank Dr. Daniel Gutmore whom I have known and admired for a number of years. I would be remiss to fail in thanking Dr. Theodore Creighton and Dr. Rikki Hatfield for their great work in editing and cleaning up my dissertation.

I thank my sister Nancy and her husband Salvatore Valente. The founder of the Italian-American Library at Seton Hall University, my brother-in-law Sal played a key role in encouraging me back to academia to earn my doctoral degree at SHU. Also, I want to acknowledge my deceased father, Armando, and my deceased sister, Linda. In some way I know that they are rejoicing over my accomplishment in God's heavenly kingdom.

This dissertation is dedicated to my mother, Leona, who is 90 years old and fortunately still with us here on this earthly plane. She has always stood especially proud of any accomplishment I have had from athletic fields, to musical venues, to my career as an educator, and to my role as her son. Mom, this dissertation is completed as a small token (I owe you many) of my appreciation for being my mom. I love you. Thank you for all you mean to me.

Here I am now, being what I might have been.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter I

Introduction

Background

For decades, there has been increasing emphasis on the use of standardized testing to evaluate student learning, gauge the quality of K-12 education, and its use as a lever for education policy initiatives. The nationwide movement to the Common Core State Standards (CCSS) in 2010 brought about another wave of standardized testing with the advent of two national testing consortia. The claimed effectiveness of standardized tests to improve student achievement was coupled with the claim that the CCSS increase the rigor of student learning. Supporters of such measures made statements such as "the standards are designed to prepare our students for college and careers by emphasizing high-level skills needed for tomorrow's world" (State of New Jersey Department of Education, 2017, p. 1).

Subsequent to a 2015 name change, the CCSS became known as the New Jersey Student Learning Standards (NJSLS). These standards appear to be inextricably tied to standardized tests in New Jersey. In fact, claims are made that the standards and the tests they produce are as the definitive tools to transmit and assess content related to college and career preparation, as they purport to focus on higher-order thinking defined as critical thinking, analysis, and problem-solving in the almost 600 public school districts across the state.

New Jersey is also home to 252 parochial Catholic private schools that use other standardized assessments to assess student learning, especially in the area of higher-order thinking. In this study, I will investigate claims made by two commercially prepared standardized tests adopted by one large New Jersey diocese. Specifically, I seek to explain how the language found in the questions on the TerraNova English/Language Arts Practice Test

(http://classroom.jc-schools.net/8th/) and the IOWA English/Language Arts Practice Test (https://www.ixl.com/standards/iowa/ela/grade-8) compare with the language that promotes higher-order thinking found in research literature.

The investigation is inspired by the fact that I serve as a principal of a Catholic pre-kindergarten through grade eight elementary school within a diocese that recently switched from the long-used TerraNova standardized testing instrument to the IOWA standardized testing instrument. The diocesan curriculum is based on the CCSS, and the diocese embraces the idea of the importance of higher-order thinking. From an instructional leadership perspective, it would be helpful to determine if either of the two assessment instruments includes language that assesses students' higher-order thinking skills with problems/questions that require complex thinking.

With regard to the IOWA Standardized Assessment, the developers claim that the test contains measures aligned to broad content standards and is reflective of a level of cognitive complexity appropriate for that child's stage of development. The developmental appropriateness is guided by research and practice in the achievement domains – the major domains of the Common Core State Standards for English Language Arts (Welch Dunbar, & Rickets, 2016). Tienken and Orlich (2013) stress the importance of setting standards and achievement targets based on cognitive development. CTB/McGraw-Hill advertises that the TerraNova test offers multiple assessments to measure important higher-order thinking skills as well as basic and applied skills. The TerraNova test is designed to generate norm-referenced achievement scores, criterion-referenced objective mastery scores, and performance-level information (Data Recognition Corporation, 2017).

Problem Statement

Controversy over the effectiveness of the CCSS and corresponding standardized assessments is not a recent phenomenon. Those who challenge standardization see acronyms such as NCLB, AYP, NAEP, and now more recently in New Jersey ASK and PAARC as monikers for movements of standardization that promote nothing more than standardized assessment instruments of questionable quality which yield equally questionable data. From Dewey (1916) to Tanner and Tanner (2007) onward, and from standardized reform efforts throughout the history of public schools in the U.S., empirical research tells us standardization is a flat-out failure in terms of increasing complex thinking such as problem solving, creative thinking, and strategic thinking (Sforza, Tienken, & Kim, 2016).

Conversely, proponents of standardization promote the positives, and its impact on curriculum and assessment. Phelps et al. (2005) deems standardization and standardized assessment as both necessary and productive. He states, that "standardized tests may be imperfect but they increase the likelihood that the consumer's aims will be served and reduce the likelihood that reports of student, teacher, and school performance will be colored by educator self-interest." Further, he claims that "from the standpoint of education's providers, standardized tests are undesirable because they restrict the ability of schools to set their own goals and priorities, and interpret outcomes accordingly. Standardized tests assess student performance in light of announced objectives, and they do so in a uniform and objective manner. The portfolio assessments preferred by educators assess student work products selected ad hoc and evaluated in a relatively subjective manner" (Phelps, et al., 2005, xvi).

The question of standardized/high-stakes testing serving to assess higher-order thinking skills with embedded complex thinking requirements is a pertinent one – and worthy of

investigation. As aforementioned, both Houghton Mifflin Harcourt and CTB/McGraw Hill claim their instruments – the IOWA and TerraNova tests respectively – challenge students with problems of higher-order thinking and cognitive complexity.

Despite claims that standardized assessment instruments challenge students to utilize complex thinking skills, little is known about how the language of the questions found on the IOWA and TerraNova English/Language Arts test for grade 8 compared to the language of higher level thinking found in empirical literature. This study is deemed significant in that educational practitioners – teachers and administrators – will be able to make decisions as to the effectiveness of the TerraNova and IOWA Standardized Assessments as per their advertising that higher-order thinking challenges are offered in their instruments vis-à-vis embedded complex thinking requirements.

Purpose of this Study

My purpose for this convergent, parallel mixed-methods study was to describe the way(s) in which the language found in the questions on TerraNova and IOWA Practice Tests in English/Language Arts for grade 8 compare with the language that promotes higher-order thinking found in research literature.

Research Questions

This study was grounded by an overarching research question: What are the types of thinking assessed by the questions in the TerraNova and IOWA English/Language Arts Practice Tests for grade 8?

1. In what way(s) does the language found in the questions on the TerraNova English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in research literature?

2. In what way(s) does the language found in the questions on the IOWA English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in research literature?
3. What differences and similarities exist in the language of complex thinking between the TerraNova and IOWA English/Language Arts Practice Tests for grade 8?

Research Design and Methodology

I conducted a convergent, parallel mixed-methods study with qualitative and quantitative content analysis methods to (a) describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 compares with the language associated with higher-order thinking found in research literature and (b) to describe and compare the complex thinking requirements found on the TerraNova and the IOWA English/Language Arts Practice Tests for grade 8. The qualitative method consisted of a content analysis of the language of the questions on each test, deductive coding, and categorizing the cognitive level of each question based on Webb's Depth of Knowledge. The quantitative aspect of the study consisted of calculating the percentages of questions categorized in each level of Webb's Depth of Knowledge.

I used Webb's Depth of Knowledge framework to categorize the language for each question of the assessments. The study used two analysts in coding the language of the questions on each test and then compared their categorizations to increase inter-rater reliability (Merriam, 2009, p. 216). To increase the coders' reliability, a read behind consensus model was used based on the work of Miles, Huberman, and Saldana (2014, p. 84). Both analysts were trained utilizing the Webb training manual (2005) on how to code each test question. The analysts used

Mayring's Step Model with protocols to guide their work. Reliability coefficients were calculated using item-total correlation.

Francis (2017) posited that Webb's DOK is an effective and useful tool for teaching and learning and for higher order thinking. Webb's DOK categorizes the levels of thinking students are expected to demonstrate which include all the taxonomies that categorize higher order thinking, and it marks and measures higher order thinking. Karen Hess (2013) noted that the Depth of Knowledge (DOK) is one of the key tools educators need to employ. The tool assists educators in better analyzing the cognitive expectation demanded by the standards, curricular activities, and assessment tasks. This study is limited to: test items contained in the 8th grade English Language Arts sections of the TerraNova and Iowa Practice Tests; coding using Webb's DOK; the time frame of the research; and the determinations derived from the data collected in the document analysis.

Webb's DOK is a hierarchy demarcated with four levels of cognitive complexity. The Wisconsin Center for Education Products and Services, Inc. (2017) notes thinking requirements as per the Webb's DOK as follows:

- Level 1 (ELA): Requires only surface understanding of text, often verbatim recall
- Level 2 (ELA): Requires processing beyond recall and observation; requires both comprehension and subsequent processing of text or portions of text; involves ordering, classifying text as well as identifying patterns, relationships, and main points
- Level 3 (ELA): Requires students to go beyond text; requires students to explain, generalize, and connect ideas; involves deep inferencing, prediction, elaboration, and summary; requires students to support positions using prior knowledge and evidence and manipulate themes across passages

- Level 4 (ELA): Requires complexity at least at the level of DOK 3, but also an extended time to complete the task, such as conducting a research project over many weeks; a project that requires extended time, but repetitive or lower-DOK tasks are not at Level 4; may require generating hypotheses and performing complex analyses and connections among texts (WCEPS, 2017).

Conceptual Framework

Webb (1997) developed a process and criteria for systematically analyzing the alignment between standards and standardized assessments. The framework also includes content consistency and that is the portion of the framework used for this study. The existing body of work on Webb's DOK offers support for the model to be employed to analyze the cognitive expectation demanded by standards, curricular activities, and assessment tasks. The model is based upon the assumption that curricular elements may all be categorized based upon the cognitive demands required to produce an acceptable response (Webb, 1997). Each grouping of tasks reflects a different level of cognitive expectation, or depth of knowledge, required to complete the task. It should be noted that the term knowledge, as it is used here, is intended to broadly encompass all forms of knowledge (i.e. procedural, declarative, etc.). Table 1 reflects an adapted version of the model (Mississippi State University, 2009).

Table 1

Webb's Depth of Knowledge (DOK) Levels

DOK Level	Title of Level
1	Recall and Reproduction
2	Skills and Concepts
3	Short-term Strategic Thinking
4	Extended Thinking

Table 1. Webb's DOK model, adapted from Mississippi State University (2009)

Webb's Depth of Knowledge (DOK) provides a respected framework with correlation to cognitive complexity. Sforza, Tienken, and Kim (2016) deem Webb's DOK to be a way to define and categorize cognitive complexity of curriculum standards and tasks. Webb's DOK definitions noted in the Research Design and Methodology section of this paper will provide the criteria base for the coding of items in the TerraNova and IOWA Tests in English Language Arts and Mathematics for grade 8. Webb's DOK alignment tool training manual (2005) provided a reference guide for the coders participating in this research study.

According to Gibbs (2007), coding is the process of organizing and sorting data. Codes serve as a way to label, compile, and organize data. They also allow one to summarize and synthesize what is happening relative to the data. In linking data collection with interpreting data, coding becomes the basis for developing the analysis. It is generally understood, then, that coding is analysis. A preassigned coding system is a system whereby "coding categories may be more or less assigned" (Bogdan & Biklen, 2016). To the extent that the Webb's DOK will be the instrument by which my mixed-methods comparison of the complex thinking requirements

embedded in the TerraNova and IOWA English Language Arts Practice Tests for grade 8 will be coded, this study is based on a preassigned coding system.

Niebling (2012) conducted a study – *Determining the Cognitive Complexity of the Iowa Core in Literacy and Mathematics: Implications and Applications for Curriculum Alignment* – for the IOWA Department of Education (as noted, this study investigates the IOWA Standardized Test). The study used Webb's DOK – with a recommendation for the DOK data to be programmed into I-CAT – in coding the cognitive complexity as per the Iowa State Department of Education's Core Standards and sub-standards. The data gleaned from this study revealed Literacy Standards as coded DOK Levels 2 and 3.

Sizemore (2015) utilized Webb's DOK in researching the *Intentional Depth of Knowledge and its Effects on K-12 Student Engagement*. As this study sought to establish the value of gathering rigor data to determine it as a component part in improving the classroom learning rigor, the DOK was used "to determine the connection between cognitive complexities as interpreted through increases in DOK". Aungst (2014) also utilized Webb's DOK in a study - *Using Webb's Depth of Knowledge to Increase Rigor*.

As noted above, the research design for this study will be a qualitative study with qualitative and quantitative content analysis methods. "The purpose of content analysis is to organize and elicit meaning from the data collected and draw realistic conclusions from it" (Bengtsson, 2016, p. 8). "As a research method, it represents a systematic and objective means of describing and quantifying phenomena. A prerequisite for successful content analysis is that data can be reduced to concepts that describe the research phenomenon by creating categories, concepts, a model, conceptual system, or conceptual map" (Elo et al., 2014, p. 1).

Webb's Depth of Knowledge/Alignment Tool is recognized nationally and internationally as a framework for study of standards and assessments as well as the alignment of the two (Burns, 2017). With its high standard of recognition, Webb's DOK is the instrument I chose to determine and compare the dissemination/occurrence of complex thinking requirements included in the TerraNova and IOWA English Language Arts (ELA) Practice Tests for grade 8.

According to Webb (1997), Depth of Knowledge encompasses multiple dimensions of thinking, including the "level of cognitive complexity of information students should be expected to know, how well they should be able to transfer the knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas" (Sforza, Tienken, & Kim, 2016, p. 6). Appendix A depicts Webb's DOK levels in English/Language Arts (Reading and Writing) as per the Web Alignment Tool (WAT, 2005) Training Manual.

Bloom's Taxonomy

Armstrong (2018) presents a background of Bloom's Taxonomy. In 1956, Benjamin Bloom with collaborators Max Englehart, Edward Furst, Walter Hill, and David Krathwohl published a framework for categorizing educational goals: *Taxonomy of Educational Objectives*. Familiarly known as Bloom's Taxonomy, it has been used by K-12 teachers since its inception. The framework elaborated by Bloom and his collaborators consisted of six major categories: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. The categories after Knowledge were presented as skills and abilities, with the understanding that knowledge was the necessary precondition for putting these skills and abilities into practice. While each category contained subcategories, all lying along a continuum from simple to

complex and concrete to abstract, the taxonomy is popularly remembered according to the six main categories from lower-order to higher-order thinking requirements (Bloom, et al., 2005):

- Create: Recall facts and basic concepts - define, duplicate, list, memorize, repeat, state
- Understand: Explain ideas or concepts - classify, describe, discuss, explain, identify, locate, recognize, report, select, translate
- Apply: Use information in new situations - execute, implement, solve, use, demonstrate, interpret, operate, schedule, sketch
- Analyze: Draw connections among ideas - differentiate, organize, relate, compare, contrast, distinguish, examine, experiment, question, test
- Evaluate: Justify a stand or decision - appraise, argue, defend, judge, select, support, value, critique, weigh
- Create: Produce new or original work - design, assemble, construct, conjecture, develop, formulate, author, investigate

Webb's DOK and Bloom's Taxonomy are – at times – superimposed in studies of complex thinking requirements, higher-order thinking, and cognitive complexity. Hess, Carlock, Jones, and Walkup (2009) noted that the superimposing of Webb's DOK and Bloom's Taxonomy was first depicted as a matrix by Hess (2005) to be used in states where the conversation about and connection of cognitive complexity as part of test design and item development processes had just begun.

Dr. Karin Hess combined Bloom's taxonomy and Webb's Depth of Knowledge into a single chart which she calls a Cognitive Rigor Matrix (Hess, Carlock, Jones, & Walkup 2009). Webb's DOK provides a comparison of varying levels or depths of knowledge applied to [academic] understanding and practices by students. Generally speaking, rigor increases as you

go from DOK 1 (lowest) to DOK 4 (highest): DOK 1 - Recall and Reproduction; DOK 2 - Basic Application of Skills and Concepts; DOK 3 - Strategic Thinking; and DOK 4 - Extended Thinking. Using the Webb's DOK systems, coders analyzed items in the TerraNova and IOWA English Language Arts Practice for grade 8.

Upon completion of the coding of items in the sections of the tests, occurrence/dissemination of items were determined and presented as data in percentage format.

Significance of the Study

For the purpose of its relevance for practical application in the field of education, this study will provide administrators, teachers, and the general public with information on the complex thinking requirements embedded in the aforementioned standardized practice tests. Implications of the study will serve to provide useful information to schools – both private and public – for curriculum development, scope and sequencing, pacing, teacher delivery of curriculum and instruction/lesson planning, student placement, and choices of nationally recognized standardized testing instruments to be administered to students.

Limitations of the Study

This study was limited to a document analysis study of the cognitive complexity of the language of the questions found on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8. The study does not cover any other versions – as per the links aforementioned – or years (both assessments copyrighted 2020). Another limitation is the coders' categorization of the questions on the tests. Coding categories may differ between coders. This suggests utilization of two types of reliability: consistency between researchers or inter-rater reliability (Haswell & Elliot, 2019), and a read behind consensus model. Both models were used in this study.

This study was additionally limited to the sample size - the number of questions published for the English Language Arts sections of the TerraNova and IOWA Practice Tests for grade 8 in the aforementioned versions. This study was also limited to the aforementioned research design and methodology employed. No other designs or methodologies were used. The results of this study were evaluated using Webb's Depth of Knowledge framework with consideration to Hess's Cognitive Rigor Matrix. No other evaluation schemes were utilized. There is a paucity of research literature regarding higher-order thinking and cognitive complexity embedded in questions on the TerraNova and IOWA standardized practice tests.

Delimitations

An initial delimitation of this study was my choice to analyze only ELA questions of the TerraNova and IOWA Practice Tests for grade 8. Math, Science, and Social Studies questions on the TerraNova and IOWA Practice Tests in grade 8 were not analyzed, nor were questions from other grade levels. Thus, the results cannot be generalized to other grade levels or other tests. Also, the results of this study were evaluated using Webb's Depth of Knowledge; therefore, the results could be different, if analyzed with a different framework.

As an additional delimitation, this study was conducted using only the assessments analyzed, the research design and methodology described, and the information on higher-order thinking found in a review of literature related to this study's topic and research questions. No other means of conducting this study were employed. A further delimitation is the time frame in which the study was conducted.

Definition of Terms

Cognitive complexity: the state or quality of a thought process that involves numerous constructs, with many interrelationships among them. Such processing is often experienced as difficult or effortful (American Psychological Association, 2018).

Creative thinking: The ability to make or do something new that is also useful or valued by others (Gardner, 1993).

Higher-order thinking resists precise forms of definition (Resnick, 1987). According to Geertsen (2003), higher-order thinking is a systematic way of using the mind to confirm existing information or to search for new information using various degrees of abstraction.

Problem solving: The identification and application of knowledge and skills that result in goal attainment (Snowman & McCown, 2015).

Standardized assessment: Snowman and McCown (2015) state that standardized tests are typically used as summative assessments (assessments of learning). They are given to students in a very consistent manner; meaning that the questions on the test are all the same, the time given to each student is the same, and the way in which the test is scored is the same for all students (Burrows, 2016).

Strategic thinking: The student uses reasoning and develops a plan or sequence of steps; the process has some complexity (Webb, 2005).

Webb's Depth of Knowledge: As Sforza, Tienken, and Kim (2016) state, "[Webb's] DOK is a way to define and categorize cognitive complexity of curriculum standards and tasks."

Organization of Dissertation

In Chapter II, the literature review situated the study in the context of previous research and scholarly material pertaining to higher-order thinking. In Chapter II, I have also presented a synthesis of literature related to this study's topic, justifies how the study addressed a gap or problem in the literature, and outlined the theoretical or conceptual framework of the study. In Chapter III, the methodology utilized in this study is presented. In Chapter IV, I organized and reported the study's main findings, including the presentation of relevant qualitative and quantitative data. Finally, in Chapter V, I provided a summary, an overview of findings, and a conclusion, as well as recommendations for future research as it relates to policy and practice.

Chapter II

Literature Review

My purpose for this convergent, parallel mixed-methods study was to describe the way(s) in which the language found in the questions on the 2016 TerraNova and the 2017 IOWA English/Language Arts and Mathematics tests for grade 8 compared with the language associated with higher-order thinking found in research literature. The purpose of this literature review was to critique the existing literature regarding higher-order thinking and higher-order thinking pertaining to requirements of standardized assessment instruments overall, specifically as it pertained to middle school standardized assessments for students in grade 8. The literature review also served to identify empirical studies about higher-order thinking and cognitive complexity of English Language Arts and Mathematics items contained in the TerraNova and IOWA tests for grade 8. Further, the literature review provided a review of relevant theories of higher-order thinking/cognitive complexity and its significance to curriculum, 21st Century Learning, Bloom's Taxonomy, Webb's DOK, the Hess Cognitive Rigor Matrix, and standardized testing. The literature review also presented a review of definitions of higher-order thinking in school curriculum. The literature review additionally identified claims of complex thinking as advertised by CTB/McGraw Hill and Houghton Mifflin Harcourt relative to the TerraNova and IOWA standardized tests respectively.

Both the TerraNova and the IOWA standardized tests are used more often in private and parochial schools than in public schools. Some public schools use the tests to assess student achievement in grades K through 2. The TerraNova is also used in some Department of Defense Dependents Schools and – as part of the California Achievement Test (CAT/6) – the state of California utilizes the TerraNova standardized test. Only the state of Iowa uses the IOWA standardized test for federal compliance. Data derived from the TerraNova and IOWA tests are used

in private and parochial schools, and in some public schools, to determine students at risk; placement in gifted and talented programs; and high school acceptance.

Literature Search Procedures

A number of sources were utilized for this peer-reviewed and non-peer-reviewed collection of literature. This included online searches (e.g. SAGE, Google Scholar, the State of New Jersey Department of Education Website, ERIC, and the Seton Hall University Databases). Searches were also conducted using key words such as higher-order thinking, cognitive complexity, standardized assessment, Webb's Depth of Knowledge (DOK), Common Core State Standards (CCSS), New Jersey Student Learning Standards (NJSLS) and complex thinking. In the literature review, I have made an effort to follow a step-by-step guide for writing a literature review as prescribed by Galvan (2006).

Overview of Current Literature

The developers of the IOWA Standardized Assessment claim that the test contains measures aligned to broad content standards and is reflective of a level of cognitive complexity appropriate for that child's stage of development. The developmental appropriateness is guided by research and practice in the achievement domains - the major domains of the Common Core State Standards - English Language Arts and Mathematics (Welch & Dunbar, 2014). Tienken and Orlich (2013) stress the importance of setting standards and achievement targets based on cognitive development. CTB/McGraw-Hill advertises that the TerraNova test offers multiple assessments to measure important higher-order thinking skills as well as basic and applied skills. The TerraNova test is designed to generate norm-referenced achievement scores, criterion-referenced objective mastery scores, and performance-level information (Data Recognition Corporation, 2017).

In that a definition of higher-order thinking resists precise forms of definition, combining the attempts of two noted researchers, the following definition can be derived. According to the attempts

of Resnick (1987) and Geertsen (2003), higher-order thinking can be defined as a systematic way of using the mind to confirm existing information, or to search for new information using various degrees of abstraction. Concisely, an exact definition of higher-order thinking does not exist. In the first phase of my search for studies, articles, journals, and dissertations pertaining to higher-order thinking, research on critical thinking became a recurring theme.

The second phase of the literature review focused on higher-order thinking language/questions embedded in the TerraNova and IOWA English/Language Arts and Mathematics tests for grade 8 and how the language/questions therein may be associated with higher-order thinking found in the literature as per Webb's DOK, Bloom's Taxonomy, and Hess's Cognitive Rigor Matrix.

With regard to the IOWA Standardized Assessment, the developers claim that the test contains measures aligned to broad content standards and is reflective of a level of cognitive complexity appropriate for that child's stage of development. The developmental appropriateness is guided by research and practice in the achievement domains - the major domains of the Common Core State Standards - English Language Arts and Mathematics (Welch and Dunbar, 2014). Tienken and Orlich (2013) stress the importance of setting standards and achievement targets based on cognitive development. Data Recognition Corporation (2017) indicated that the TerraNova assessment requires higher-order thinking skills and that the assessment measures a great depth of knowledge. A further claim is that the reliability of the test is demonstrated by the high internal consistency coefficients and the standard errors associated with the scores.

Investigation into Webb's Depth of Knowledge (DOK) with consideration to Hess's Cognitive Rigor Matrix and Bloom's Taxonomy constituted a third phase of the literature review. With its high standard of recognition, Webb's DOK is the instrument I have chosen to determine and compare the dissemination/occurrence of complex thinking requirements included in the IOWA and TerraNova

English Language Arts (ELA) and Mathematics Tests for grade 8. According to Webb (1997), Depth of Knowledge encompasses multiple dimensions of thinking, including the "level of cognitive complexity of information students should be expected to know, how well they should be able to transfer the knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas" (Sforza, Tienken, & Kim, 2016, p.6).

In searching the topics of complex and higher-order thinking, a surfeit of peer-reviewed literature can be found. In much of the literature (Allen & Wern, 2017; Barrington & Casner-Lotto 2006; Tan & Halili, 2015) there is a recurring message that there is a need for complex thinkers and problem solvers in the workplace. Public perception and various authors have highlighted the importance of schools with regard to its responsibility in providing the nation's workforce with the complex thinkers/problem solvers it deems necessary for productivity. Hoffman (2015) posited that "work provides powerful opportunities to learn, and the workplace is where many young people are most receptive to applying academic skills and content as well as using critical interpersonal and intrapersonal capacities—i.e., the collection of knowledge and skills referred to as 'deeper learning'" (p. 2).

Literature in the area of higher-order thinking – for many years – highlights a number of benefits students derive from instruction that promotes such cognitive complexity. Eisenman and Payne (1997) note that studies have reported that thinking skills instruction improve the attitudes of students toward school (Dalton & Goodrum, 1991), reading comprehension (Payne & Manning, 1992), and locus of control (Eriksson, 1990). Allen, Connor and De Castro (2017) purported that "higher-order/critical thinking [goes] beyond the mere accumulation of data but rather engages [one] in meditative and independent thinking. It facilitates the clear evaluation of data and allows a critical

thinker to draw sensible and accurate conclusions from filtered information that generally results in effective problem solving and decision-making" (p. 17). Barak, Ben-Chaim, and Uri (2007) state that the ever-changing and challenging world necessitates that students go beyond the building of their knowledge capacity as they need to develop their higher-order thinking skills, such as critical systems thinking, decision making, and problem solving . This is evident in various studies which state that research has indicated the need for improving critical thinking skills among students since many of them fail to utilize sophisticated reasoning even at the college level (Halpern, 1998; Kuhn, 1999).

High-stakes testing, standardized curriculum, and 21st Century skills are keywords imperative to education practitioners. Yen and Halili (2015) posited that "in the 21st century, one critical aspect in discussing effective teaching and learning is examining the effectiveness of teachers in developing students' capability to think while ensuring content mastery at the same time. The role of teachers in inculcating [higher-order thinking] HOTS is [an] important aspect of teaching HOTS effectively" (p. 41). With the advent of Common Core State Standards (CCSS) – now the NJSLS – and its impact on modes of mandated standardized testing, education practitioners are accountable for ensuring students score well on the administered standardized testing instruments. Accountability assessments in the State of New Jersey include: Partnership for Assessment of Readiness for College and Careers (PARCC), TerraNova, IOWA, and MAP Growth, etc.

The CCSS – now the NJSLS – have served for approximately a decade as a standards template nationwide. The Common Core State Standards website makes a claim that the CCSS are informed by the highest, most effective standards from states across the U.S. and countries around the world. A claim is also posed that the standards define the knowledge and skills students should gain throughout their K-12 education in order to graduate high school prepared to succeed in entry-level careers, introductory academic college courses, and workforce training programs

(http://www.corestandards.org/about-the-standards/development-process/). Further, it is claimed that the standards are research and evidence-based; clear, understandable, and consistent; aligned with college and career expectations; based on rigorous content and application of knowledge through higher-order thinking skills; built upon the strengths and lessons of current state standards; informed by other top performing countries in order to prepare all students for success in our global economy and society.

Brookhart (2010) divides definitions of higher-order thinking into three categories: (1) those that define 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 (as cited in Collins, 2014). The critical thinking category includes definitions that refer to "reasonable, reflective thinking that is focused on deciding what to believe or do" (p. 3-4). Although transfer is difficult to measure with a standardized test, some tests claim to measure critical thinking and problem solving.

Literature Inclusion Criteria

Research used in this review included:

A. Dissertations on higher-order thinking

B. Peer-reviewed studies focused on key terms such as higher-order thinking, strategic thinking, cognitive complexity, creative thinking, and problem solving

C. Mainstream, non-peer reviewed literature/information on the topics of higher-order thinking, standardized assessment, standards, Webb's Depth of Knowledge (DOK), Bloom's Taxonomy, and Hess's Cognitive Rigor Matrix.

D. Peer and non-peer reviewed articles on the Common Core State Standards (CCSS) and the New Jersey Student Learning Standards (NJSLS) in middle school English/Language Arts and Mathematics

E. Government reports

F. Classic literature on the subject of higher-order thinking

Methodological Issues with Existing Literature

Templier and Paré (2015) list six steps essential to formulating a literature review: formulating the research question(s) and objective(s), searching the existing literature, screening for inclusion, assessing the quality of primary studies, extracting data, and analyzing data. I have found the following steps as presenting issues methodologically. With regard to formulating the research questions, there will be interpretive comparison regarding the language of higher-order thinking embedded in the TerraNova and IOWA English/Language Arts tests and how it is associated with the language of higher-order thinking found in research literature. Coding as per Webb's DOK will serve to clarify any ambiguities.

Paré, Tate, Johnstone, and Kitsiou (2016) posited that - in a literature review - an effort should be made to be as comprehensive as possible in order to ensure that all relevant studies - peer-reviewed and non-peer-reviewed - are included in the literature review. This, they note, will ensure that conclusions are based on this all-inclusive knowledge base. This literature review will be limited in its scope as there is a scarcity of relevant studies about the way(s) in which the language found in the questions on the TerraNova and the IOWA English/Language Arts and Mathematics tests for grade 8 compares with the language associated with higher-order thinking found in research literature.

As aforementioned, lacking is a formal definition of higher-order thinking. Often, literature refers to such terms as critical thinking and/or problem solving as being synonymous with higher-order thinking. One may ascertain that the authors of literature on the topics of critical thinking and/or problem solving assume that their audience will be those in the field of education, and thus suppose that the terminology used can be interpreted as such. In place of specific definition, much literature on

higher-order thinking provides examples of student performance tasks that exemplify higher-order thinking. Additionally, literature searches on higher-order thinking yield "to critical thinking and problem solving as essential components of higher-order thinking skill set development" (Sydoruk, 2018, p. 17).

Search engines also provided scare results relative to the term cognitive complexity. Higher education and some hits in the field of nursing were found. The results found in the search are not applicable; therefore not useful to my study as this study focused on middle school, specifically grade 8. Further, searches on higher-order thinking, critical thinking, cognitive complexity, and problem solving yielded a paucity of information when these terms were applied to searches intending to link them to the IOWA and TerraNova standardized testing instruments.

Also, the frameworks utilized in this study can be nebulous in defining the higher-order thinking that is investigated in this study. Bloom's Taxonomy does provide some insight on cognitive processing and thinking, but the Taxonomy can be seen as one-dimensional in its focus on the verbs associated with the domains. In 2001, Anderson and Krathwohl, et al, presented a structure for rethinking Bloom's Taxonomy. Whereas the original taxonomy possessed one dimension, the revised taxonomy table applied two dimensions: cognitive processes and knowledge. The cognitive processes resemble those found in the original taxonomy, but placement on the taxonomy continuum has changed slightly (e.g. evaluation no longer resides at the highest level) and descriptions have been expanded and better differentiated for analyzing educational objectives (Anderson & Krathwohl et al., 2001).

21st -Century Skills/Learning

21st-Century Learning and Skills have been recently coined to describe how schools should prepare students in order to ensure that they are workforce ready for the future. Non-peer-reviewed

literature views 21st-century learning largely as a future workforce necessity. Voogt and Roblin (2012) state that "knowledge has become vital in the 21st -century and people need to acquire such skills to enter the workforce. In general, 21st-century skills include collaboration, communication, digital literacy, citizenship, problem solving, critical thinking, creativity and productivity" (as cited in van Laar, van Deursen, van Dijk, & de Haan, 2017, p. 577). These skills are labeled 21st-century skills to indicate that they are more related to current economic and social developments than related with those of the past century characterized as an industrial mode of production. Fadel (2008) posited that 21st-century skills cover the areas of critical thinking, problem-solving, life, and career skills with an emphasis on innovative/creative thinking as well as information, media, and technology skills.

The P21 Framework for 21st Century Learning was developed with input from a coalition of educators, education experts, and business leaders who collaborated to define and illustrate the skills, knowledge, expertise, and support systems that students need to succeed in work, life, and citizenship. The Framework continues to be used by thousands of educators and hundreds of schools in the U.S. and abroad to put 21st century skills at the center of learning. All elements of the Framework are critical to ensure 21st century readiness for every student (Partnership for 21st-Century Skills, 2007a, 2016). In terms of learning and innovation skills and competencies, the P21 framework states the following: "Learning and innovation skills are what separate students who are prepared for increasingly complex life and work environments in today's world and those who are not. They include: creativity and innovation; critical thinking and problem solving; communication; and collaboration [the 4C's]" (p. 2).

Peer reviewed literature on 21st-century skills/learning reveals themes similar to the non-peer-reviewed literature on the topic. Boholano (2017) stated that "education in the 21st century highlights globalization and internationalization" (p. 21). McCoog (2008) and Zaidieh (2012) noted that it is

essential that 21st-century learners possess both self-direction and an ability to collaborate with individuals and groups – as in social networking. The National Research Council initiated an investigation into the topic of teaching and assessing 21st century skills, beginning in 2005. Out of this work came a framework for categorizing the types of knowledge and skills students need for college and career readiness: (1) cognitive skills, including critical thinking, non-routine problem solving, and systems thinking; (2) interpersonal skills, including complex communication, social skills, teamwork, cultural sensitivity, and dealing with diversity; and (3) intrapersonal skills, including self-management, time management, self-development, self-regulation, adaptability, and executive functioning (National Research Council, 2011).

A recurrent theme in 21st-century skills/learning literature – both peer and non-peer reviewed literature – is that the 21st-century differs from the 20th-century relative to the capabilities people need for work, citizenship, and self-actualization. Peer-reviewed literature about 21st-century skills indicates that 21st-century skills are different from 20th-century skills primarily due to the emergence of very sophisticated information and communications technologies (Dede, 2009). Economists Frank Levy and Richard Murnane (2004) highlighted a crucial component of what constitutes 21st-century knowledge and skills:

> Declining portions of the labor force are engaged in jobs that consist primarily of routine cognitive work and routine manual labor—the types of tasks that are easiest to program computers to do. Growing proportions of the nation's labor force are engaged in jobs that emphasize expert thinking or complex communication—tasks that computers cannot do. (p. 53–54)

Proponents of 21st-Century learning note that ability to find and analyze information from multiple sources and then utilize the information in decision making and idea creation are critical to

workers in nearly all sectors of the economy (Murnane & Levy, 2004). Even if 21st-century skills are not new – they have perhaps taken on a new importance. Critics of 21st-Century learning claim that teaching such skills detracts from the teaching of core content and may serve to "water-down standards and weaken teaching…they argue these types of higher order skills cannot be measured in a reliable, cost-effective, or scalable ways" (Silva, 2009, p. 360).

Overall, proponents of 21st-Century skills/learning espouse that these skills will prepare the next generation for work and life. They proclaim that students must learn to think for themselves while in the field, solve complex, often multifaceted problems, and be able to communicate their decisions effectively. Silva (2009) noted that proponents of 21st-century skills and learning highlight a need for a new type of workforce reality that requires a next generation of college students and workers who are independent thinkers, problem solvers, and decision makers. As a proponent of 21st- century skills, Crockett (2015) indicated that students need to be able to think and work creatively in both digital and non-digital environments to develop unique and useful solutions. In addition, he notes that students must be able to think analytically, which includes proficiency with comparing, contrasting, evaluating, synthesizing, and applying without instruction or supervision.

Critics of 21st Century skills/learning argue against the push for these skills. They call 21st-century skills a meaningless term and that these skills are diverting educators from the more important work of teaching core content. There is nothing new about these skills, they say, and emphasizing them will water down standards and weaken teaching (Silva, 2009). Kohn (2009) questioned just what the phrase 21st-Century skills means. Kohn wondered if it really can be known just what specific skills will be needed in the future. Other critics noted that 21st-century skills/learning were an effort by technology companies – which make up the bulk of the group's membership – to gain more influence over the classroom, and that it is an attempt to gain profit by selling more equipment to

schools (Sawchuk, 2009). Critics also claim that the notion of 21st-Century skills is nebulous, and they question whether or not 21st-century skills/learning requires skills vastly different than any other century's skills.

Higher-Order Thinking

In searching the topics of complex and higher-order thinking, a surfeit of peer-reviewed literature can be found. In much of the literature (Allen, 2017; Barrington & Casner-Lotto, 2006; Tan, 2015) there is a recurring message that there is a need for complex thinkers and problem solvers in the workplace. Public perception and various authors have highlighted the importance of schools with regard to its responsibility is providing the nation's workforce with the complex thinkers/problem solvers it deems necessary for productivity. For example, Hoffman (2015) posited that work provides powerful opportunities to learn, and the workplace is where many young people are most receptive to applying academic skills and content as well as using critical interpersonal and intrapersonal capacities—e.g., the collection of knowledge and skills referred to as deeper learning.

Rehmat (2015) indicated that while we move forward in the 21st-century, higher-order thinking skills – especially as they pertain to the disciplines of math and science achievement – are requisite skills necessary to meet the educational requirement of Science Technology Engineering and Math (STEM) careers. Educators need to think of innovative ways to engage and prepare students for current and future challenges while cultivating an interest among students in STEM disciplines. An instructional pedagogy that can capture students' attention, support interdisciplinary STEM practices, and foster higher-order thinking skills is problem-based learning. Problem-based learning embedded in the social constructivist view of teaching and learning (Savery & Duffy, 1995) promotes self-regulated learning that is enhanced through exploration, cooperative social activity, and discourse (Fosnot, 1996).

For many years, literature in the area of higher-order thinking highlights a number of benefits students derive from instruction that promotes such cognitive complexity. Eisenman & Payne (1997) noted that research studies have indicated that thinking skills instruction can improve students' attitudes toward school (Dalton & Goodrum, 1991), reading comprehension (Payne & Manning, 1992), and locus of control (Eriksson, 1990). Allen, Connor and De Castro (2017) purported that higher-order/critical thinking goes beyond one just accumulating data, but rather engages one in meditative and independent thinking. This type of thinking – they believe – facilitates the clear evaluation of data and allows a critical thinker to draw sensible and accurate conclusions from filtered information that generally results in effective problem solving and decision-making. "Our ever-changing and challenging world requires students, our future citizens, to go beyond the building of their knowledge capacity; they need to develop their higher-order thinking skills, such as critical systems thinking, decision making, and problem solving" (Barak, Ben-Chaim, & Zoller, 2007, p. 354). Various studies and research has indicated the need for improving critical thinking skills among students since many of them fail to utilize sophisticated reasoning even at the college level (Halpern, 1998; Kuhn, 1999).

One of the leading challenges that the nation's future workforce faces is the use of critical thinking skills in the workplace. Employees will need the ability to use information from a broader and more impartial perspective in order to make more informed decisions and also see a comprehensive view of any situation (Ferri-Reed, 2013). The U.S. Department of Labor (2000) identified critical thinking and higher-order thinking skills as a raw material for some vital workplace skills, including problem-solving and decision-making. It is generally known that the workforce of the future will require employees whose work will involve creativity, problem-solving, and – among other aptitudes – critical analysis (Ananiadou & Claro, 2009; Rimini & Spiezia, 2016).

Individuals have to process complex information, think systematically and take decisions weighing different forms of evidence. They also have to continuously update their skills to match the rapid changes in technology at the workplace. More fundamentally, in order to seize the new opportunities that digital technologies are opening in many areas, individuals have to develop the right set of skills to make a meaningful use of these technologies.

Smith (1992) and Zohar and Dori (2003) exclaim that higher order thinking skills (HOTS) is a thinking process, which consists of complicated procedures and needs to be based on various skills such as analysis, synthesis, comparison, inference, interpretation, assessment, and inductive and deductive reasoning to be employed to solve unfamiliar problems. Educators have an assortment of HOTS strategies that include several concepts. Krulik and Rudnick (1993) state that HOTS includes 1) recall thinking, 2) basic thinking, 3) critical thinking, and 4) creative thinking. Byrnes (1996) classifies HOTS into 4 levels; 1) the application level, 2) the analysis level, 3) the synthesis level, and 4) the evaluation level. Anderson and Krathwohl (2001) support the concepts of Bloom's Taxonomy Revised (more on Bloom's Taxonomy later in this chapter), and classify cognitive approaches to learning into six levels; 1) remembering, 2) understanding, 3) applying, 4) analyzing, 5) evaluating, and 6) creating (as cited in Budsankom et al, 2015).

Schultz and FitzPatrick (2016) posited that higher-order thinking includes reasoning, judgement, and critical thinking. It is the incorporation of the three that leads to one's ability to solve problems. Demand for higher-order or complex thinking does not necessarily refer to problem solving or task difficulty. Rather, it refers to the level or type of thinking required for task completion.

As noted, gleaning the perfect definition of higher-order thinking is difficult. Newmann (1988) defines it simply as he indicated that higher-order thinking "signifies the challenge and expanded use of the mind" (p. 59). Brookhart (2010), Keefe et al. (1992), Paul and Elder (2006), and others present

myriad representations of higher-order thinking, yet without one specific definition. Offered more often than not, one finds in the literature scenario form examples on what higher-order thinking looks like. Higher-order thinking is also associated with synonyms such as critical thinking, creative thinking, metacognitive thinking, complex thinking, etc.

With the advent of 21st-Century Skills and its focus on the type of thinking needed for one to succeed in the 21st-century, one might be deceived into assuming that higher-order thinking is strictly a 21st-century phenomenon and/or skill. In fact, higher-order thinking has been around the teaching/learning venue for a good part of three centuries in American education. Though never mentioning the term higher-order thinking as such, Dewey (1902) purported that students should be active learners, examining and inquiring in order to hone advanced thinking skills. Curiosity and creativity – according to Tanner (2016) – are the beginning stages of higher-order thinking. He noted that curiosity and creativity help children learn via exploration and interactions.

Further, in connecting the need for 21st-century skills with the need for a prepared future workforce, economists note that "expert thinking [involves] effective pattern matching based on detailed knowledge; and metacognition, the set of skills used by the stumped expert to decide when to give up on one strategy and what to try next" (Levy & Murnane, 2004, p. 75). The need for 21st-century skills to be embedded in the delivery of curriculum and instruction in the classroom is essential to the formation of the 21st-century workforce. According to Johnson (2009), adopting a 21st-century curriculum should blend knowledge, thinking, innovation skills, media, Information and Communication Technology (ICT) literacy, and real life experience in the context of core academic subjects. In order to achieve authentic learning that is demanded in the 21st-century, students must engage in the learning environment effectively and develop 21st-century skills such as critical thinking, problem solving, and collaboration. In this way, students will be prepared with the necessary

knowledge and life skills that will help them be successful in their future careers (Lombardi, 2007). Curriculum in the 21st century should focus on the construction of knowledge and encourage students to produce the information that has value or meaning to them in order to develop new skills. Preparing curriculum to be connected with the real world can support student participation, their motivation and understanding for the academic subjects, as well as preparing them for adult life (Lombardi, 2007). Common Core State Standards (CCSS) researchers stated, "The standards are designed to be robust and relevant to the real world, reflecting the knowledge and skills that our young people will need for success in college and careers and be positioned to compete successfully in the global economy" (Common Core State Standards Initiative [CCSSI], 2010, p.6).

Research on 21st-century learning explores both character traits and higher-order thinking aptitudes. Ball, Joyce, and Anderson-Butcher (2016) found in their research of middle school students that they held moderately high perceptions in the areas of leadership, responsibility, working with others effectively, and adaptability. Collins (2014) identified research that highlighted the importance of teaching higher-order thinking skills to prepare young men and women to live in the 21st-century. While learning for recall requires thinking, the higher-order thinking is in transfer. That is, students not only acquire the knowledge and skills, but also can apply them to new situations.

Higher-order thinking skills include critical thinking, problem solving, decision making, and creative thinking (Lewis & Smith, 1993). According to Dewey (1933), thinking does not happen automatically, it must rather be brought to mind by way of problems and questions or by some disconcertion or perplexity. He noted the term reflection as a form of higher-order thinking and proposed that:

> Demand for the solution of a perplexity is the steadying and guiding factor in the entire process of reflection. Where there is no question of a problem to be solved or a difficulty to be

> surmounted, the course of suggestions flows on at random; we have the first type of thought described. If the stream of suggestions is controlled simply by their emotional congruity, their fitting agreeably into a single picture or story, we have the second type. But a question to be answered, an ambiguity to be resolved, sets up an end and holds the current of ideas to a definite channel. Every suggested conclusion is tested by its reference to this regulating end, by its pertinence to the problem in hand. (Dewey, 1910, p. 4)

In light of his writing about the demand for a solution of a perplexity, Dewey further stated:

> We may recapitulate by saying that the origin of thinking is some perplexity, confusion, or doubt. Thinking is not a case of spontaneous combustion; it does not occur just on "general principles." There is something specific which occasions and evokes it. General appeals to a child (or to a grown-up) to think, irrespective of the existence in his own experience of some difficulty that troubles him and disturbs his equilibrium, are as futile as advice to lift himself by his boot-straps. (Dewey, 1910, p. 4)

Kauchak and Eggen (1998) stated that it is important to teach students to think about their own thinking processes. Higher-order thinking skills include critical, logical, reflective, metacognitive, and creative thinking. According to King, Rohani, and Goodson (1997), the ability of higher-order thinking skills are activated when individuals encounter unfamiliar problems, uncertainties, questions, or dilemmas. When these skills are nurtured and well developed, one can perform better during explanations and making decisions as well as grow their intellectual skills. Development of higher order thinking skills relies on their lower level thinking skills thus making higher order thinking skills grounded with lower level thinking skills. To be able to think critically, prior knowledge of subject matter content is necessary. According to King et al. (1997) as well, appropriate teaching strategies

and learning environments facilitate a learner's growth as does student persistence, self-monitoring, and open-minded, flexible attitudes.

Higher-order thinking, though difficult to specifically define, does find some specificity in terms of defining and/or describing higher-order thinking questions. Higher-order questions are those that the students cannot answer by simply recalling information or by reading the information verbatim from the text. Higher-order questions put advanced cognitive demand on students. They encourage students to think beyond literal questions. Higher-order questions promote critical thinking skills because these types of questions expect students to apply, analyze, synthesize, and evaluate information instead of simply recalling facts.

> For instance, application questions require students to transfer knowledge learned in one context to another; analysis questions expect students to break the whole into component parts such as analyze mood, setting, characters, express opinions, make inferences, and draw conclusions; synthesis *questions* have students use old ideas to create new ones using information from a variety of sources; and evaluation *questions* require students to make judgments, explain reasons for judgments, compare and contrast information, and develop reasoning using evidence from the text. (Bogdanovich, 2014; p. 1)

Tankersley (2005) purported that higher-order thinking requires students to evaluate, synthesize, analyze, and interpret. With regard to assisting students to develop higher-order thinking skills, Alvermann and Phelps (1998) tell us, "The curriculum must expand to include information and activities that explicitly support students in learning to think well. The emphasis is less on the mastery of information measured by a recall-based assessment and more on learning how to use one's mind well, to synthesize and analyze skillfully" (p. 69).

Bloom's Taxonomy and Higher-Order Thinking Skills (HOTS)

In 1956, Benjamin Bloom with collaborators Max Englehart, Edward Furst, Walter Hill, and David Krathwohl published a framework for categorizing educational goals: *Taxonomy of Educational Objectives*. Familiarly known as Bloom's Taxonomy, this framework has been applied by generations of K-12 teachers and college instructors in their teaching. The framework elaborated by Bloom and his collaborators consisted of six major categories: *Knowledge, Comprehension, Application, Analysis, Synthesis*, and *Evaluation*.

The categories after Knowledge were presented as skills and abilities, with the understanding that knowledge was the necessary precondition for putting these skills and abilities into practice. While each category contained subcategories, all lying along a continuum from simple to complex and concrete to abstract, the taxonomy is popularly remembered according to the six main categories.

Bloom's taxonomy is taught in a majority of teacher-education programs in the United States. As such, it may be among the most well-known educational theories among teachers nationally. As Watson (2019) noted:

> While Bloom's Taxonomy is not the only framework for teaching thinking, it is the most widely used, and subsequent frameworks tend to be closely linked to Bloom's work.... Bloom's aim was to promote higher forms of thinking in education, such as analyzing and evaluating, rather than just teaching students to remember facts - rote learning. (p. 1)

Bloom's taxonomy was designed with six levels to promote higher-order thinking. The six-levels were: knowledge, comprehension, application, analysis, synthesis, and evaluation. The taxonomy's levels were later revised as remembering, understanding, applying, analyzing, revising, and

creating. The lower-order thinking skills (LOTS) involve memorization, while higher-order thinking requires understanding and applying that knowledge.

The top three levels of Bloom's taxonomy - which is often displayed as a pyramid, with ascending levels of thinking at the top of the structure - are analysis, synthesis, and evaluation. These levels of the taxonomy all involve critical or higher-order thinking. Students who are able to think are those who can apply the knowledge and skills they have learned to new contexts. Looking at each level demonstrates how higher-order thinking is applied in education.

Analysis

Analysis, the fourth level of Bloom's pyramid, involves students using their own judgment to begin analyzing the knowledge they have learned. At this point, they begin understanding the underlying structure of knowledge and also are able to distinguish between fact and opinion. Some examples of analysis would be:

- Analyze each statement to decide whether it is fact or opinion.
- Compare and contrast the beliefs of W.E.B. DuBois and Booker T. Washington.
- Apply the rule of 70 to determine how quickly your money will double at 6 percent.
- Illustrate the differences between the American alligator and the Nile crocodile.

Synthesis

Synthesis, the fifth level of Bloom’s taxonomy pyramid, requires students to infer relationships among sources such as essays, articles, works of fiction, lectures by instructors, and even personal observations. For example, a student might infer a relationship between what has been read in a newspaper or article and what the student has observed. The high-level/higher-order thinking of synthesis is evident when students put the parts or information they have reviewed together to create new meaning or a new structure. At the synthesis level, students move beyond relying on previously

learned information or analyzing items that the teacher is giving to them. Examples of questions in an educational of assessment setting that would involve the synthesis level of higher-order thinking might include:

- What alternative would you suggest for ___?
- What changes would you make to revise___?
- What could you invent to solve___?

Evaluation

Evaluation, the top level of Bloom's taxonomy, involves students making judgments about the value of ideas, items, and materials. Evaluation is the top level of Bloom's taxonomy pyramid because it is at this level that students are expected to mentally assemble all they have learned to make informed and sound evaluations of the material. Some questions involving evaluation might be:

- Evaluate the Bill of Rights and determine which is the least necessary for a free society.
- Attend a local play and write a critique of the actor's performance.
- Visit an art museum and offer suggestions on ways to improve a specific exhibit. (Watson, 2019)

Cultivating Higher-Order Thinking

Sternberg (2005) stated that the goal of education is to promote higher-order thinking by teaching for successful intelligence. He proposed that education needs to capitalize on an individual's strengths while working toward improvement of their weaknesses through analytical, creative, and practical instruction. Following this Triarchic Theory of Intelligence, Sternberg believed that students could be provided with skills and abilities for higher level thinking and real life success. Higher-order thinking is sometimes associated with terms such as critical or creative thinking. Paul and Elder (2007) posited that critical thinking/higher-order thinking is the foundation of a strong education. Citing Bloom's Taxonomy of Thinking Skills, they believed that the goal in education is to move

students from lower to higher-order thinking; from knowledge (information gathering); to comprehension (confirming); to application (making use of knowledge); to analysis (taking information apart); to evaluation (judging the outcome); to synthesis (putting information together) and creative generation.

Cultivating higher-order thinking skills is important in that not to do so can have deleterious impact with regard to the development of such skills. One such negative impact is known as functional fixedness. Psychological studies on creative problem solving have explored factors that determine whether or not one's knowledge about the world or experience with a particular kind of problem or situation can facilitate efforts to solve a new problem with similar features. The concept of analogical transfer - the transfer of a basic structure acquired through one or more instances to another instance. This structure is sometimes called a paradigm. The analogical transfer phenomenon has been well-established in the creativity literature (e.g., Gick & Holyoak, 1980, 1983; Holyoak, 1984, 2005). Analogical transfer is not always positive. Under certain circumstances, prior knowledge or experience with a particular example or solution strategy may have negative effects for creative thought (e.g., Gentner, 1983; Osman, 2008). Functional fixedness or fixation is an instance of such negative transfer, wherein a solver's experience with a particular function of an object impedes using the object in a novel/different way during creative problem solving (Duncker, 1945; Scheerer, 1963).

Functional fixedness, which is studied in the field of cognitive psychology, originated in Duncker's (1935) seminal study of how adults solved various mathematical and practical problems. The study was published in his book *Psychologie des produktiven Denkens*. Oelze (2020) noted that Duncker argued that while functional fixedness is a necessary perceptive and cognitive skill, it can hamper problem solving and creativity. Seifert (2009) noted that – despite Duncker's claim that functional fixedness is on some level a necessary perceptual and cognitive skill – he deemed it one of

the two common problems that hinder problem solving. He posited that an individual can get so used to one particular purpose for an object that other uses may be overlooked. German & Barrett (2005) define functional fixedness "as a tendency to regard the functions of objects and ideas as fixed" (from Seifert, 2009; p. 190). Seifert (2009) expanded the concept by indicating that functional fixedness may also be referred to as a response set, "the tendency for a person to frame or think about each problem in a series in the same way as the previous problem, even when doing so is not appropriate to later problems" (p. 190). He added that functional fixedness and the response set are obstacles in problem representation, the way that a person understands and organizes information provided in a problem. "If information is misunderstood or used inappropriately, then mistakes are likely - if indeed [a] problem can be solved at all" (p. 190).

As noted earlier, the "level of cognitive complexity of information students should be expected to know, how well they should be able to transfer the knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas" (Sforza, Tienken, & Kim, 2016, p.6). In the category of transfer, Anderson, Krathwohl et al (2001) define transfer in how it differs from retention. They note that two of the most important educational goals are (1) to promote retention and (2) to promote transfer which, when it occurs, indicates meaningful learning. Retention requires that students remember what they have learned, whereas transfer requires students not only to remember but also to make sense of and be able to use what they have learned.

While learning for recall requires thinking, the higher-order thinking is in transfer. When transfer happens, students not only acquire the knowledge and skills, but also can apply them to new situations. It is this kind of thinking, according to Brookhart (2010) that applies to life outside of school where thinking is characterized by a series of transfer opportunities rather than as a series of

recall tasks to be executed. The critical thinking category includes definitions that refer to reasonable, reflective thinking that is focused on deciding what to believe or do (Norris & Ennis, 1989), and artful thinking, which includes reasoning, questioning and investigating, observing and describing, comparing and connecting, finding complexity, and exploring viewpoints (Barahal, 2008).

Hmelo and Ferrari (1997) indicated that it is vital to assist learners in the development of higher-order thinking skills as well as a flexible knowledge base. They suggested that cognitive science and education studies/research suggests that both aims can be attained by having students learn through solving problems. With its emphasis on both strategies and content, problem-based learning (PBL) - they noted - is specifically well suited to achieving the aforementioned aims. In PBL, student learning begins with a problem to be solved and includes cycles of reflection on the problem-solving experience. Zhong (2002) found that "critical thinking is a prudent and doubtful thinking activity. It guides people's beliefs and actions through observing, reflecting, reasoning and communicating psychological processes" (from Zhou, 2018, p. 349-350). Wen (2009) proposed that critical thinking is conscious thinking through certain criteria of thinking. Ultimately, this leads to one making rational judgments, and then improving rational thinking. He proposed that reflective thinking is the thinking skill.

Critical/higher-order thinking is based on constructivist theory (also referred to as structuralism) – a branch of cognitive psychology. Constructivists Jean Piaget, Jerome Bruner, and George Kelly are the three recognized authorities in the field of cognitive psychology. An important concept of constructivism is schema, which refers to the way individuals understand and think about the world. It can also be regarded as a framework or organizational structure of mental activity. Schema is the starting point and core of cognitive structure or it is the basis of human cognition. "Schema formation

and change is the essence of cognitive development, cognitive development is influenced by three processes: assimilation, adaptation and balance" (Zhou, 2018, p. 350)

Common Core State Standards (CCSS)

Common Core State Standards (CCSS), released in 2010 (now the NJSLS), were championed by state leaders, including governors and state commissioners of education from forty-eight states, two territories and the District of Columbia, through their membership in the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO). In the State of New Jersey, the State Board of Education adopted the Common Core State Standards in lieu of the formerly implemented Core Curriculum Content Standards. The Common Core State Standards lay a pre-designed framework for precisely what kindergarten through high school students should both know, and be able to do. Despite the fact that the Common Core State Standards are not advertised as a curriculum, and that school districts are held responsible for development district curricula, the Common Core State Standards are a key driving force in the development of individual district curricula. The CCSS influence curriculum development and standardized assessment.

Proponents of the Common Core State Standards initiative – namely state school chiefs and governors – espoused the value of consistent, real-world learning goals and launched this effort to ensure all students, regardless of where they live, are graduating high school prepared for college, career, and life (Common Core State Standards Initiative, 2018). Revolving around two key and specific tenets, the CCSS were conceived and divided into the nomenclatures of first – College and Career Readiness Standards; and second – the K-12 Standards.

On the topic of complex thinking requirements – as per the research questions posed in this study – Hiebert and Mesmer (2013), caution those who may believe in the complex thinking/cognitive

complexity requirements that the Common Core State Standard proponents claim exist. In their research on the CCSS English Language Arts Standards, they have found, "standard 10 addresses text complexity using almost the same language for all grade levels, while interchanging the relevant grade-level bands and specifying the degree of proficiency required within different grades of a band" (p. 45). Tienken and Zhao (2010) offered that teachers bound to adherence to the CCSS are forced to follow programmed or scripted programs that do not create learning; and that merely imitate processes that ultimately result in the loss of vital skills and learning experiences for students. Thus, teachers teaching and assessing to promote higher-order thinking skills, creativity, cognitive complexity, and complex thinking requirements are diminished because of the CCSS.

The standardized Common Core State Standards and their influence on standardized testing, and standardization in general, receive further criticism from oppositional forces. Proponents advocate that United States students lag behind in standardized test ranking when compared to students in such countries Finland or cities such as Singapore. Nationally, advocates of standardization embrace the palette and canvas of global economic competitiveness and international test scores (especially as per math and science) to paint a dour portrait of America's future. Tienken (2013) posits that there is no empirical evidence that demonstrates a correlation between the test rankings of U.S. students to any indicators of economic prowess. Yet, global economic competitiveness and the ranking of U.S. students on international assessments continue to be utilized as arguments standardization promoters proclaim as a support for reform efforts such as the Common Core State Standards (CCSS).

Proponents submit that standardization – and curriculum innovation such as the Common Core State Standards – will promote creativity and innovation (higher-order thinking skills). Such pundits of standardization cite international test scores and note national security and economic competitiveness and growth is at risk. Tienken (2017) posited that "[t]he focus of curriculum policy and leadership

should be to facilitate the democratic curricula processes that relate to the Curriculum Paradigm and that address the complementary functions of public schools" (p. 130).

Standardized Testing

Concisely, standardized tests are often referred to as assessments. These standardized or high stakes tests provide schools, districts, and states with information about student achievement (or lack thereof). Standardized tests can also determine whether or not students are promoted to the next grade level, and they can be used to determine special placement of students - e.g. gifted and talented, or referral for evaluation to the child study team for potential special education classification and programming. (Great Schools Partnership, 2014) noted that when tests are standardized, a situation is created whereby a particular group of students – with varied academic aptitudes – take the same test that will be scored and analyzed the same way. The score result for each student is then compared to the rest of the group to see how well the students performed. Mukeredzi, Bertram, and Christiansen (2018) show support of standardized testing as they noted assessment as a key component of learning as it helps students learn because students are able to see how they are doing in a class; thus, they are able to determine whether or not they understand course material. Assessment can also help motivate students. If students know they are doing poorly, they may begin to work harder.

Proponents of standardized testing or assessment exclaim various virtues such as it provides useful data to help guide the delivery of curriculum and instruction; to determine appropriateness of curriculum and pacing guides; for Response to Intervention (RTI) plans for students in need; and for student placement to name a few. "Educators, parents, and the public depend on accurate, valid, reliable, and timely information about student academic performance. The availability of test data is important in that it can "improve instruction, identify the needs of individual students, implement targeted interventions, and help all students reach high levels of achievement" (U.S. Department of

Education, 2013, p. 1). In studying the correlation of CCSS to standardized testing, Porter, McMaken, Hwang, and Yang (2011) propose through their research that there exists a focus on the standards but an absence of a coherent assessment.

However, those who oppose standardization – and the standardized testing it promotes – will argue that it ignores variables such as the all-important stages of cognitive development (as per Piaget or Vygotsky). They further note that what reform standardization advocates claim will benefit our nation's schools and its students is based on little more than political ideology. Tienken and Orlich (2013) stress the importance of setting standards and achievement targets based not on political ideology, but rather on cognitive development. Jouriles (2014) states that standardized tests are not necessary because they rarely show what students do not already know. He stated, "Ask any teacher, and she can tell you which students can read and write…that telling usually comes in the form of letter grades, or evaluations that break down progress on skill" (p. 1).

McNeil (2000) offered that proponents of standards believe that these will ensure quality education – particularly since they would allow for testable results that would enable comparisons to be made between schools and across districts. Those who oppose standards and standardization find that they are a bureaucratic and politically-driven phenomenon with derivations from within the organizational structures of schooling, rather than from the theories of child development and learning. They further claim that standardization has traditionally signaled separateness from teaching, learning, and curriculum.

Both the TerraNova Test and the IOWA Test are nationally norm-referenced assessments. **"**Norm-referenced refers to standardized tests that are designed to compare and rank test takers in relation to one another. Norm-referenced tests report whether test takers performed better or worse than a hypothetical average student, which is determined by comparing scores against the performance

results of a statistically selected group of test takers, typically of the same age or grade level, who have already taken the exam" (Great Schools Partnership, 2015, https://www.edglossary.org/norm-referenced-test/). As an example, the publisher of the Iowa Test of Basic Skills will collect data from thousands of Iowa students to create norms, so that the new form will also be a normed test or normed instrument.

A gap of limitation to note is the paucity of literature with regard to qualitative studies with qualitative and quantitative content analysis methods to determine and compare the complex thinking requirements of the questions found on the TerraNova English Language Arts and the IOWA English Language Arts Tests for grade 8. Carter (2015) conducted a quantitative study with the purpose of investigating if reading to students for five days a week, for 30 minutes per day, could significantly impact TerraNova Reading scores. Karam, Stecher, Tsai, Grimm, and Schweig (2016) chose a primarily qualitative case-study approach to study and examine how various educational stakeholders in a school system were implementing the 2009 DoDEA Mathematics Standards and the steps they were taking to align their mathematics program with the CCR standards. For this, the TerraNova Test was utilized as a measure. McCutchen, Jones, Carbonneau, and Meuller (2016), utilized the IOWA Standardized Test to conduct a qualitative (with quantitative content analysis methods) longitudinal study of the relationship between students' mindsets and their standardized test performance. Hand, Norton-Meier, Gunel and Akkus (2014) used a 3-year mixed-methods research design examining the embedding of language and argumentation into elementary science classrooms. In this research, student IOWA standardized testing results were utilized as one data tool.

The Data Recognition Corporation (DRC) which vends the TerraNova Test for reading, language, math, science, and social studies, espouses that the assessment features a carefully researched test blueprint that reflects today's challenging standards for content and performance

(DRC, 2015). It further claims to do this via customized research studies, studies conducted by the DRC. The IOWA Test – recently renamed the IOWA Test of Educational Development (ITED) – claims that rather than testing a student's content knowledge, the ITED endeavors to evaluate students' skills in a variety of areas, especially based on problem solving and critical analysis of texts. Perusing the Literature that the Data Recognition Corporation (DRC) puts forth for the TerraNova assessment, and the literature that Seton Testing Services and the University of Iowa puts forth for the IOWA test, a reader would believe that both standardized assessments challenge students with higher-order thinking items and are full of complex thinking requirements.

Background on the TerraNova Standardized Test

Linda W. Thompson (2013), a former National Assessment Consultant and current professional development consultant for CTB/McGraw-Hill, claims that the TerraNova standardized assessment meets the gold standard in test development and scoring. CTB/McGraw-Hill is renowned for over eighty-five years of innovation and research excellence, which ensures high validity and reliability required for a credible assessment. She further posits that TerraNova 3 content was developed to reflect national standards so that every school, regardless of geographic location or specific curriculum, can compare how its students' learning measures up to the achievement of students across the nation.

CTB/McGraw-Hill notes that TerraNova assessment items reflect the rigor expected of student learning today - 21st-century learning - and that all TerraNova assessment items are classified according to Rankin-Hughes Dimensions of Thinking, a framework similar to Bloom's taxonomy. In addition it is claimed that, all reading, language, and math items are aligned with

the CTB Depth of Knowledge (DOK) framework, and that Reading, Language, and Mathematics items in the TerraNova assessment reflect one of four DOKs:

1. Recognize and Recall
2. Use Fundamental Concepts and Procedures
3. Conclude and Explain
4. Evaluate, Extend, and Make Connections

CTB/McGraw-Hill asserts that analyzing student score report data in the objectives reports, TerraNova test content and DOKs can be used to inform instruction. It is professed that educators can identify objectives on which students perform well and those where students are challenged. By using data gleaned from TerraNova testing, educators can target specific areas for reinforcing classroom instruction. Further, the developers of the TerraNova note that student score data information can also be used to target students in need of intervention in specific areas. Combined with classroom work and assessments, it is proposed that the TerraNova standardized assessment objectives data provide a useful start at identifying groups for differentiated instruction.

To put its product in a historical perspective, the test company that produces the TerraNova assessment highlights that CTB/McGraw-Hill has an 85-year record of innovation and excellence in assessment, and serves more than 18 million students in all 50 states and in 49 countries. My research found that the company stands as one of the first American publishers to introduce objective, standardized achievement tests to schools. CTB/McGraw-Hill to be a leader in testing with recognized products for online interim and formative, adult, and language proficiency assessment. They profess to advance the use of student performance data to inform instructional decision making. CTB/McGraw-Hill also provides technologies that include web-

based assessment and reporting, student response device software, and artificial intelligence for automated scoring of student essays. CTB/McGraw-Hill is part of McGraw-Hill, a division of The McGraw-Hill Companies (NYSE: MHP). McGraw-Hill espouses that they are a leading global provider of instructional, assessment and reference solutions that empower professionals and students of all ages (McGraw-Hill Communications Team, 2012).

Background on the IOWA Standardized Test

In 1929, the University of Iowa initiated the nation's first major state-wide testing program for high school students. Directed by E.F. Lindquist – considered one of the fathers of standardized achievement testing – the Iowa program had several remarkable features: every school in the State of Iowa could participate on a voluntary basis; every pupil in participating schools was tested in key subjects; new editions of the achievement tests were published annually; and procedures for administering and scoring tests were highly structured. Results were used to evaluate both students and schools, and schools with the highest composite achievement received awards. In addition, Lindquist was among the first to extend the range of student abilities tested. The Iowa Tests of Basic Skills and the Iowa Test of Educational Development became tools for diagnosis and guidance in grades three to eight and in high school, respectively. The IOWA test has had an influence on swinging the pendulum of educational testing in the direction of diagnosis and monitoring, and away from classification and selection (U.S. Congress, Office of Technology Assessment, 1992).

Since its beginnings, the IOWA test has grown as a nationally used standardized assessment in schools. The Iowa Assessment is a group-administered achievement test for students in grades K through 12. The test is designed to measure a student's knowledge in subject areas that students have learned in school – reading, language arts, mathematics, science and social studies. The Iowa

Assessments were released in the 2011-2012 school year and have replaced the Iowa Test of Basic Skills (ITBS) and the Iowa Test of Educational Development. The developers claim that the newest exam format has been redesigned to better align the exam with the Common Core Standards and that it is better aligned with other state standardized tests such as the Smarter Balanced Assessment exam. The Iowa Assessments are used by school districts nation-wide to assess a student's college and career readiness. They are also becoming more prevalent as a component of determining whether a student will be admitted into a schools Talented and Gifted Program (TAG or GATE program).

The developers of the IOWA assessment claim that the test contains measures aligned to broad content standards and is reflective of a level of cognitive complexity appropriate for that child's stage of development. The developmental appropriateness is guided by research and practice in the achievement domains - the major domains of the Common Core State Standards - English Language Arts and Mathematics (Welch & Dunbar, 2014). Tienken and Orlich (2013) stress the importance of setting standards and achievement targets based on cognitive development.

Webb's Depth of Knowledge (DOK)

Webb (1997) developed a process and criteria for systematically analyzing the alignment between standards and standardized assessments. His framework also includes content consistency and that is the portion of the framework used for this study. The existing body of work on Webb's DOK offers support for the model to be employed to analyze the cognitive expectation demanded by standards, curricular activities, and assessment tasks (Webb, 1997). "The model is based upon the assumption that curricular elements may all be categorized based upon the cognitive demands required to produce an acceptable response. Each grouping of tasks reflects a different level of cognitive expectation, or depth of knowledge, required to complete the task. It should be noted that the term knowledge, as it is used here, is intended to broadly encompass all forms of knowledge – i.e.

procedural, declarative, etc."

(https://www.aps.edu/sapr/documents/resources/Webbs_DOK_Guide.pdf, 2009, p. 5).

Depth of Knowledge includes multiple dimensions of thinking which includes the level of cognitive complexity of information students should be expected to know, how well they should be able to transfer the knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas" (Webb, 1997, p.15). I used Webb's (1997; 2007) four DOK levels as the guide to coding the questions on the TerraNova and IOWA English Language Arts and Mathematics tests for grade 8.

Depth of knowledge considers a number of forms of knowledge such as based on facts knowledge - declarative, and practical know-how - procedural knowledge. These two forms of knowledge are linked; thus, they are the base that structures creative and strategic thinking opportunities. Levels 1 and 2 of Webb's DOK center on declarative and procedural knowledge (recall, reconstruction). Levels 4 and 4 of Webb's DOK include creative and strategic thinking. These levels afford students opportunities for deeper, analytical, and divergent types of thinking (Sforza, Tienken, and Kim, 2015). Amongst researchers, it is argued that creativity and strategic - higher-order thinking - occur at DOK Levels 3 and 4.

Webb's Depth of Knowledge (DOK) provides a respected framework with correlation to cognitive complexity. Sforza, Tienken, and Kim (2016) deem Webb's DOK to be a way to define and categorize cognitive complexity of curriculum standards and tasks. The DOK is employed in a number of states and school districts to determine the depth of assessment, state, and curriculum standards vis-a-vis their correlation to higher-order thinking/cognitive complexity. In that Webb's DOK is utilized in this study to categorize how the language found in the questions on the TerraNova IOWA English/Language Arts Practice Tests for grade 8 compare with the language associated with

higher-order thinking found in the research literature, it is also important in assisting teachers plan and deliver curriculum and instruction that promotes a greater range of cognitive demand upon students.

Karen Hess (2013) noted that Norman Webb's Depth-of-Knowledge (DOK) schema has become one of the key tools educators can employ to analyze the cognitive demand (complexity) intended by the standards, curricular activities, and assessment tasks. Webb (1997) developed a process and criteria for systematically analyzing the alignment between standards and test items in standardized assessments. Since then, the process and criteria have demonstrated application to reviewing curricular alignment as well. The model categorizes assessment tasks by different levels of cognitive expectation, or depth of knowledge, required to successfully complete the task. Hess (2004) further articulated the model with content specific descriptions for use by classroom teachers and organizations conducting alignment studies.

Norman Webb's alignment system is one of several alignment systems developed in relatively recent time (Webb, 1997). In this alignment system, Webb outlays five criteria for alignment between standards and assessments. The five criteria are:

- Categorical Concurrence: This criterion measures the extent to which the same or consistent categories of content appear in the standards and the assessments. The criterion is met for a given standard if there are more than five assessment items targeting that standard.
- Depth-of-Knowledge Consistency: This criterion measures the degree to which the knowledge elicited from students on the assessment is as complex within the context area as what students are expected to know and do as stated in the standards. The criterion is met if more than half of targeted objectives are hit by items of the appropriate complexity.
- Range-of-Knowledge Correspondence: This criterion determines whether the span of knowledge expected of students on the basis of a standard corresponds to the span of

knowledge that students need in order to correctly answer the corresponding assessment items/activities. The criterion is met for a given standard if more than half of the objectives that fall under that standard are targeted by assessment items.

- Balance of Representation: This criterion measures whether objectives that fall under a specific standard are given relatively equal emphasis on the assessment.
- Source of Challenge: This criterion is met if the primary difficulty of the assessment items is significantly related to students' knowledge and skills in the content area as represented in the standards (Webb et al., 2005).

Theoretical Framework

Myriad models and frameworks exist for promoting and assessing higher-order thinking. Webb's Depth of Knowledge and Bloom's Taxonomy are often recognized as forerunners in this area. Though similar in the terminology used in varied levels of their frameworks, they are different in the area of focus pertaining to higher-order thinking. Bloom's Taxonomy, referring specifically to the knowledge-based taxonomy, concentrates on actions students perform in order to demonstrate an understanding of a particular concept. Webb's Depth of Knowledge, on the other hand, focuses on a task's level of cognitive complexity, which encompasses the number of connections a student makes, the level of reasoning, and reflective and self-monitoring processes utilized in order to effectively complete a task (Jirka & Hableton, 2005).

In that this study considered the Hess' Cognitive Rigor Matrix, information pertaining to the matrix is in order. Hess's Cognitive Rigor Matrix superimposes Webb's and Bloom's frameworks in a model that unifies the actions of Bloom's with the cognitive complexity of Webb's. This unification of the two frameworks provides educators a better understanding of how to develop higher-order thinking and complex thinking in the classroom. Overall, Hess'

Cognitive Rigor Matrix blends the actions with the cognitive complexity of tasks and activities being asked of students in order to create a holistic model that can encourage higher order thinking (Hess et al., 2009).

Strongly influencing the theoretical framework for this study is the writing of John Dewey (1933) specifically with regard to his powerful landmark work, *How We Think.* I have used Dewey's ideas and concepts and linked them to Webb's Depth of Knowledge levels as justification of the theoretical framework of this study. In his 1933 work, *How We Think*, Dewey distinguishes between four different modes of thinking: imagination, belief, stream of consciousness, and reflection.

Dewey acknowledges that imagination, belief, and stream of consciousness are certainly part of our thinking activities, yet they do not necessarily contribute to learning and even less to lifelong learning. Reflection, however, plays a different role. Dewey defines reflection as "active, persistent, and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it and the further conclusions to which it tends" (p. 9).

Reflection is active. When we reflect we examine prior beliefs and assumptions and their implications. Reflection is an intentional action. Dewey noted, "A demand for the solution of a perplexity is the steadying [and] guiding factor in the entire process of reflection" (p. 14). Dewey added, "the function of reflective thought is, therefore, to transform a situation in which there is experienced obscurity, doubt, conflict, disturbance of some sort into a situation that is clear, coherent, settled, harmonious (p. 100). Reflection starts with discomfort during an experience and leads a person to a balanced state. It takes time and focus to reach clarity of thought.

Dewey wrote that reflection "gives an individual an increased power of control" (p. 21). It "emancipates us from merely impulsive and merely routine activity…it converts action that is merely appetitive, blind and impulsive into intelligent action" (p. 17). It is not enough just to have an experience. Reflection directs that experience to learning and deeper insights.

Reflective thinking takes time and requires one to engage in several different phases or aspects of reflective thought:

1. Perplexity: responding to suggestions and ideas that appear when confronted with a problem.
2. Elaboration: referring to past experiences that are similar.
3. Hypotheses: developing several potential hypotheses.
4. Comparing hypotheses: finding some coherence within these hypotheses.
5. Taking action: experiencing "mastery satisfaction, enjoyment" when selecting and then acting on these hypotheses (Dewey, 1933, pp. 106-115).

Dewey (1933) asserts that these are not steps but aspects of reflective activity. An individual may stop at some point and find it necessary to go back and, for example, collect more experiences. A key point is that informed action follows this reflective thinking process and leads to more ideas and therefore generates more experience on which to reflect. Dewey posited that "reflective thinking impels to inquiry" (p. 7). In fact, to Dewey, reflective thinking fosters the development of three attitudes that further the "habit of thinking in a reflective way." These three attitudes are:

- Openmindedness (freedom from prejudice)
- Wholeheartedness or absorbed interest
- Responsibility in facing consequences (Dewey, 1933, p. 33)

These dispositions are the foundation for education that gives people "a personal interest in social relationships and control and the habits of mind that secures social changes without introducing disorder" (Dewey, 1944, p. 99).

In How We Think, John Dewey (1910) writes that so profuse and varied is our use of 'thinking' and 'thought' that it is not easy to define just what we mean by them. His intention in chapter one is 'to find a single consistent meaning'. It is worthwhile noting that Dewey's (1916, p.158) passions for 'thinking' and for the importance of the development of thinking skills in schools can again be found in Democracy and Education where he writes: "no one doubts, theoretically, the importance of fostering in school good habits of thinking". Dewey (1916, p.159) further argues that "thinking is the method of intelligent learning, of learning that employs and rewards mind". Thus, "thinking originates in situations where the course of thinking is an actual part of the course of events and is designed to influence the result. The object of thinking is to help reach a conclusion, to project a possible termination on the basis of what is already given...and since the situation in which thinking occurs is a doubtful one, thinking is a process of inquiry, of looking into things, of investigating" (Dewey, 1916, p.154). For Dewey then, "all thinking is research, and all research is native, original, with him who carries it on, even if everybody else in the world already is sure of what he is still looking for". In "How we Think", Dewey (1910, p.1) argues that – in the first place – thought can be used loosely to refer to anything that 'comes to mind' or that 'goes through our heads'. In this regard "to think of a thing is just to be conscious of it in any way whatsoever". In this loose sense, thinking signifies everything that is 'in our heads' or that 'goes through our minds'. Ennis (1964, p.599) calls this 'poor critical thinking'. With this, it can be said that the responsibility of educators is to shun inclinations to promote 'poor critical thinking' among our students, but instead to strive towards

'thinking' in its critical, more reflective sense, or what Ennis calls 'good critical thinking'. Schafersman (1991, p.3) argues that critical thinking is critical inquiry. To that end, critical thinkers investigate problems, ask questions, pose new answers that challenge the status quo, discover new information that can be used for good or ill, question authorities and traditional beliefs, challenge received dogmas and doctrines, and often end up possessing power in society greater than their numbers. In another publication Schafersman (1997, p.4) posited that "the only way to escape both deception by others and the far more common trait of self-deception is to repeatedly and rigorously examine your basis for holding your beliefs. You must question the truth and reliability of both the knowledge claims of others and the knowledge you already possess". For Dewey (1910, p.4), "thought denotes belief resting upon some basis, that is, real or supposed knowledge going beyond what is directly present. It is marked by acceptance or rejection of something as reasonably probable or improbable". Regarding 'reflection' Dewey (1910, p.8) noted it is unequivocal that "reflection thus implies that something is believed in (or disbelieved in), not on its own direct account, but through something else which stands as witness, evidence, proof, voucher, warrant; that is, as ground of belief". In Democracy and Education Dewey (1916, p.150) contends that 'thought' or 'reflection' "is the discernment of the relation between what we try to do and what happens in consequence. No experience having a meaning is possible without some element of thought". Dewey (1910, p.6) defines 'reflective thought' as "active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends". For Dewey (1910, p.57), "reflection is turning a topic over in various aspects and in various lights so that nothing significant about it shall be overlooked - almost as one might turn a stone over to see what its hidden side is like or what is covered by it". Dewey argues that in

speaking of reflection, “we naturally use the words weigh, ponder, and deliberate - terms implying a certain delicate and scrupulous balancing of things against one another. Closely related names are scrutiny, examination, consideration, inspection - terms which imply close and careful vision”.

Dewey (1910, p.10) picks out sub-processes which are involved in every reflective operation. These are: (a) a state of perplexity, hesitation, doubt; and (b) an act of search or investigation directed toward bringing to light further facts which serve to corroborate or to nullify the suggested belief. Perplexity forces us to carefully scrutinize what is before us; to look for evidence that will support belief in favor of either of the roads. For Dewey (1910, p.11), “thinking begins in what may fairly enough be called a forked-road situation, a situation which is ambiguous, which presents a dilemma, which proposes alternatives”. He contended that “the origin of thinking is some perplexity, confusion, or doubt. Thinking is not a case of spontaneous combustion; it does not occur just on ‘general principles’. There is something specific which occasions and evokes it” (Dewey, 1910, p.12). Perplexity, confusion, or doubt then necessitates an act of search or investigation. Dewey reminded us that the next step is suggestion of some way out the formation of some tentative plan or project, the entertaining of some theory which will account for the peculiarities in question, the consideration of some solution for the problem. Grant and Zeichner (1984, p.104) note that “much of what Dewey had to say on this matter was directed specifically at teachers and prospective teachers, and his remarks remain very relevant for those in the process of becoming teachers”.

In seeking to compare Dewey’s ideas with regard to thinking with the ideas of Webb on the same, a start can be found in the following. Dewey espoused - as aforementioned - that any thought that goes through one’s mind is thinking. Dewey was interested, therefore, in ‘how we

think' as opposed to expectations set forth for how we do so. In contrast, Webb's Depth of Knowledge has come to be associated with required or expecting thinking as per school curricula and assessment. Miller (2017) notes that to engage students in meaningful conversations and discourse around core content matter we can start to build the classroom conversation by scaffolding using Webb's (2002) Depth of Knowledge (DOK) questions. Miller (2017) noted "DOK questions refer to the level of thinking a certain question requires in formulate an answer" (p. 1). Webb's DOK has been highly heralded since many states transitioned to College and Career Readiness, Common Core, and Next Generation Science Standards, etc. Thus, the DOK has become associated with the delivery of curriculum and instruction, standardization, teacher criterion-referenced based assessment, and standardized testing as it categorizes the levels of thinking students are expected to demonstrate. Francis (2017) purported that – out of all taxonomies that categorize higher-order thinking – Webb's DOK is the most aligned to the Structure of Observed Learning Outcomes (SOLO) Taxonomy and its levels. SOLO Taxonomy Levels are: (1) Prestructural: at this level the learner is missing the point; (2) Unistructural: a response based on a single point; (3) Multistructural: a response with multiple unrelated points; (4) Relational: points presented in a logically related answer; and (5) Extended Abstract: demonstrating an abstract and deep understanding through unexpected extension.

Webb first introduced the concept of Depth of Knowledge (DOK) in a paper entitled "Research Monograph No. 8: Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education (1997). In this paper, Webb's purpose was to define criteria for judging the alignment between expectations and assessments (Webb, 1997). He further clarified that DOK consistency serves as an attribute under content focus criteria. Webb explained how DOK can vary dimensionally as follows:

- The level of cognitive complexity of information students should be expected to know;
- how well they should be able to transfer this knowledge to different contexts;
- how well they should be able to form generalizations;
- and, how much prerequisite knowledge they must have in order to grasp ideas (Webb, 1997)

Francis (2017) noted that in essence, depth of knowledge designates how deeply students must know, understand, and be aware of what they are learning in order to attain and explain answers, outcomes, results, and solutions. The DOK also designates the extent to which students are expected to use or transfer learning as applied to scholastic and real world contexts.

Francis (2017) - in contrast to Miller (2017) - noted that "the focus of [the] levels [in Webb's DOK] is not on the type of thinking or even the kind of knowledge students are expected to demonstrate. He proposed that what Webb established was "the context - the scenario, the setting, or the situation - which students will express and share the depth and extent of their learning" (p.2). For example, are students expected to acquire knowledge (DOK 1)? Apply knowledge (DOK 2)? Analyze knowledge (DOK 3)? Augment knowledge (DOK 4)?

In my research of Dewey and Webb, I see the following links between tenants regarding and levels of thinking that they addressed. As aforementioned, Dewey's work addressed how we think. Webb - despite how the DOK has become tied the delivery of curriculum and instruction, standardization, teacher criterion-referenced based assessment, standardized testing and the kind of thinking students are expected to demonstrate – I deem Miller's (2017) proposition valid. He noted – as aforementioned – that "the focus of the levels in Webb's DOK is not on the type of thinking or even the kind of knowledge students are expected to demonstrate. He proposed that what Webb established was "the context - the scenario, the setting, or the situation - which

students will express and share the depth and extent of their learning" (p.2). This indicated to me that Webb was - as was Dewey - focused on how we think.

Further, I can ascertain connections between Dewey's phases or aspects of reflective thought and levels of Webb's Depth of Knowledge. Dewey defined reflective thought as 'active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends' (Dewey 1933: 118). He set out five phases or aspects. According to Smith (1996, 1999, 2013) these states of thinking are as follows:

1. Suggestions in which the mind leaps forward to a possible solution: Perplexity.
2. An intellectualization of the difficulty or perplexity that has been felt (directly experienced) into a problem to be solved: Elaboration.
3. The use of one suggestion after another as a leading idea, or hypothesis, to initiate and guide observation and other operations in collection of factual material: Hypothesis.
4. The mental elaboration of the idea, or supposition as an idea or supposition (reasoning, in the sense in which reasoning is a part, not the whole, of inference): Comparing hypothesis.
5. Testing the hypothesis by overt or imaginative action. (See Dewey 1933: 199-209): Taking action.

Webb (2009) stated that DOK Level 1 - Recall and Reproduction involves the basic type of thinking that requires one to recall or reproduce knowledge and/or skills. One employing Level 1 thinking either knows or does not know an answer to a presented problem. This aligns with Dewey's notion of perplexity in that - even though DOK Level 1 is of basic status, one is

confronted with a problem. Of course, the perplexity phase or aspect overlaps in some way with all four DOK levels.

DOK Level 2 - Working with Skills and Concepts - was explained by Webb (2009) as requiring one to engage in some mental processing above and beyond recalling or reproducing a response or answer to a problem (perplexity). At this level, DOK Level 2 and Dewey's elaboration aspect connect. Level 2 may require one to "make use of information in a context different from the one in which it was learned" (Webb, 2009, p.9). Mental/thinking processes that mark this level include: summarize, estimate, organize, classify, and infer. One might employ reference to past experiences (elaboration) as this DOK level speaks to how knowledge can be used in solving a problem or answering a question.

Thinking associated with DOK Level 3 - Short-Term Strategic Thinking - is explained by Webb (2009) as "short-term use of higher-order thinking processes such as analysis and evaluation to solve real-world problems with predictable outcomes" (p.11). Mental processes that are associated with DOK Level 3 include, but are not limited to the following verbs: analyze, explain, generalize, create, and - as a connection to Dewey's third phase or aspect - hypothesize.

DOK Level 4 - Extended Strategic Thinking - is explained by Webb (2009) as demanding "extended use of higher-order thinking processes such as synthesis, reflection, assessment, and adjustment of plans over time" (p. 13). Mental skills/strategic thinking processes associated with this level include, but are not limited to the following verbs: synthesize, reflect, conduct, prove, report, create, appraise, design, create, etc. DOK Level 4 correlates with Dewey's fourth and fifth phases/aspects - comparing hypothesis and taking action respectively. In that aforementioned verbs associated with DOK Level 4 are action verbs, Dewey's taking action (phase/aspect five) connects with this level. Further, in that "taking action" can only be had

subsequent to one making sense (coherence) with what is learned and/or hypothesized, Dewey's phase/aspect number four connects with DOK Level 4.

The theoretical framework above guided my research, and has assisted me in determining what will be measured, and what relationships - in a mix-method study - I sought. In Chapter III, an explanation of the methodology of this study is presented. Chapter III includes detail on how I conducted a mixed-methods study with qualitative and quantitative content analysis methods to (a) describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 compares with the language associated with higher-order thinking found in research literature and (b) to describe and compare the complex thinking requirements found on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8.

Chapter III

Methodology

My purpose for this convergent, parallel mixed-methods study was to describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts (ELA) Practice Tests for grade 8 compared with the language associated with higher-order thinking found in the research literature. This study focused specifically on TerraNova and IOWA middle school standardized practice assessments for students in grade 8. Standardized test scores are widely used by high schools for student placement. Catholic schools also use standardized test scores for student acceptance as well as scholarship awards. Eighth graders – generally at the age of thirteen to fourteen years of age – fit into what Piaget (1936) described as the Formal Operational Stage of Cognitive Development. David L (2014) indicated that, in the formal operational stage (approximately age 12 and up), one's cognition reaches its final form. By this stage, a person no longer requires concrete objects to make rational judgements. He/she is now capable of deductive and hypothetical reasoning. His/her ability for abstract thinking is very similar to the abstract thinking ability of an adult. Cherry (2019) posited that – in the formal operational stage – one's ability to think about abstract ideas and situations is a key hallmark of the formal operational stage of cognitive development. The ability to systematically plan for the future and reason about hypothetical situations are also critical abilities that emerge during this stage. Concisely, grade 8 students – as per Piaget's formal operational stage of cognitive development – are ready for and able to (barring cognitive dysfunction) thinking critically and engaging in higher-order thinking challenges.

The developers of the IOWA Standardized Assessment claim that the test contains measures aligned to broad content standards and is reflective of a level of cognitive complexity

appropriate for that child's stage of development. The developmental appropriateness is guided by research and practice in the achievement domains – the major domains of the Common Core State Standards – English Language Arts and Mathematics (Welch & Dunbar, 2014). Tienken and Orlich (2013) stressed the importance of setting standards and achievement targets based on cognitive development. CTB/McGraw-Hill advertises that the TerraNova test offers multiple assessments to measure important higher-order thinking skills as well as basic and applied skills. The TerraNova test is designed to generate norm-referenced achievement scores, criterion-referenced objective mastery scores, and performance-level information.

Research Questions

This study was guided by three research questions:

This study was grounded by an overarching research question: What are the types of thinking assessed by the questions in the TerraNova IOWA English/Language Arts Practice Tests for grade 8?

1. In what way(s) does the language found in the questions on TerraNova English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in research literature?
2. In what way(s) does the language found in the questions on the IOWA English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in research literature?
3. What differences and similarities exist in the language of complex thinking between the TerraNova and IOWA English/Language Arts Practice Tests for grade 8?

Hypothesis

Based on the literature of what multiple choice test formats test well and do not test well, the hypothesis for the study is directional in nature. The language of questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 will associate more frequently with Webb's Level 1 and Level 2 lower level thinking more so than Levels 3 and 4 higher level thinking.

Research Design and Methodology

I conducted a convergent, parallel mixed-methods design with qualitative and quantitative methods to describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 compares with the language associated with higher-order thinking found in research literature; and to describe and compare the complex thinking requirements found on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8.

I used Webb's Depth of Knowledge framework to categorize the language for each question of the assessments. The study used one expert coder/analyst who worked with me in coding the language of the questions on each test. The coders compared their categorizations to increase inter-rater reliability (Merriam, 2009). To increase the coders' reliability, a double-rater read behind consensus model was used based on the work of Miles, Huberman, and Saldana (2014). Both coders/analysts were trained by Dr. Christopher H. Tienken utilizing the Webb training manual (2005) on how to code English/Language Arts questions. The analysts used Mayring's Step Model with protocols to guide their work. Reliability was attained using the double-rater read behind consensus model.

Francis (2017) posited that Webb's DOK is an effective and useful tool for teaching and learning and for categorizing higher-order thinking. Webb's DOK categorizes the levels of thinking students are expected to demonstrate which include all the taxonomies that classify higher order thinking. Depth of Knowledge (DOK) also marks and measures higher-order thinking. Karen Hess (2013) stated that the Depth of Knowledge (DOK) is one of the key tools educators need to employ. The tool assists educators in better analyzing the cognitive expectations demanded by the standards, curricular activities, and assessment tasks. This study is limited to test items contained in the 8th grade English Language Arts TerraNova and Iowa Practice Tests; coding using Webb's DOK; the time frame of the research; and the determinations derived from the data collected in the document analysis.

Webb's DOK is a hierarchy demarcated with four levels of cognitive complexity. The Wisconsin Center for Education Products and Services, Inc. (WCEPS) describe thinking requirements and levels of cognitive complexity in Webb's DOK as follows:

- Level 1 (ELA): Requires only surface understanding of text, often verbatim recall
- Level 2 (ELA): Requires processing beyond recall and observation; requires both comprehension and subsequent processing of text or portions of text; involves ordering, classifying text as well as identifying patterns, relationships, and main points
- Level 3 (ELA): Requires students to go beyond text; requires students to explain, generalize, and connect ideas; involves deep inferencing, prediction, elaboration, and summary; requires students to support positions using prior knowledge and evidence and manipulate themes across passages
- Level 4 (ELA): Requires complexity at least at the level of DOK 3, but also an extended time to complete the task, such as conducting a research project over many

weeks; a project that requires extended time, but repetitive or lower-DOK tasks are not at Level 4; may require generating hypotheses and performing complex analyses and connections among texts (WCEPS, 2017).

Webb's purpose in formulating the Depth of Knowledge scale was to define criteria for judging the alignment between expectations and assessments (Webb, 1997). He further clarified that DOK consistency serves as an attribute under content focus criteria. Webb explained how DOK can vary dimensionally as follows:

- The level of cognitive complexity of information students should be expected to know;
- how well they should be able to transfer this knowledge to different contexts;
- how well they should be able to form generalizations; and,
- how much prerequisite knowledge they must have in order to grasp ideas (Webb, 1997)

Webb's DOK (2002) was noted in his publication *Depth of Knowledge in the Four Content Areas*. He addressed how depth of knowledge could be addressed in English/Language Arts, Mathematics, Science, and Social Studies (from Francis, 2017):

- DOK-1: Recall and reproduce data, definitions, details, facts, information, and procedures. (knowledge acquisition)
- DOK-2: Use academic concepts and cognitive skills to answer questions, address problems, accomplish tasks, and analyze texts and topics. (knowledge application)
- DOK-3: Think strategically and reasonably about how and why concepts, ideas, operations, and procedures can be used to attain and explain answers, conclusions, decisions, outcomes, reasons, and results. (knowledge analysis)

- DOK-4: Think extensively about what else can be done, how else can learning be used, and how one could personally use what one has learned in different academic and real world contexts. (knowledge augmentation)

Francis (2017) noted that Webb's DOK levels establish how in-depth one will express and share learning. DOK-1 is content-specific, focusing on the specific text or topic being taught and learned. DOK-2 and DOK-3 are item-oriented, focusing on how and why learning can be transferred and used to explain reasons, relationships, and results. DOK-4 is extensive and practical, focusing on how and why learning can be transferred and used across the curriculum and beyond the classroom. Hess (2013) describes these levels not so much as steps but rather as 'ceilings' that establish how far or in-depth students will study and share knowledge and thinking.

Research Bias

Research bias could have factored into this research study. I am in my thirty-ninth year of work in education. I have served as a teacher for twelve years and as an administrator – both building-level and central office – for twenty-seven years. Bias regarding standardized testing and standardization in general could have contributed to research bias based on my professional background. Heshmat (2015) purported that confirmation bias occurs directly from the influence desire has on beliefs. For example: when people would like a certain idea or concept to be true, they tend to believe it is true. Confirmation bias could be equated with and cause one to be motivated to think wishfully. Bias error leads an individual to stop gathering information when the evidence gathered so far confirms the views or prejudices one would like to be true. Confirmation bias is the tendency to search for, interpret, favor, and recall information in a

way that affirms one's prior beliefs or hypotheses. Confirmation bias is a type of cognitive bias and a systematic error of inductive reasoning (Heshmat, 2015).

I have administered/proctored standardized tests as a teacher, served as a test coordinator, reviewed questions on myriad practice tests/teacher criterion referenced-based tests and analyzed voluminous test data as an administrator. Because of the experience I have had, I may have harbored preconceived notions and assumptions as to what I might have found in this study. I could be prone to research/confirmation bias based on my practical professional experience in the area of assessment. A bias might also exist based on my own practice as a teacher, and by way of the hundreds of observations of the delivery of curriculum and instruction I have conducted as an administrator.

My effort to minimize the chance of confirmation bias occurring was vital to this study. I was committed to remaining open-minded/objective, and that I recognized and acknowledged evidence that supported any preconceived notions I may have potentially brought to this study. Complete confirmation bias is never 100 percent achievable. In order ensure bias did not affect the study, I used Webb's Depth of Knowledge framework to categorize the language for each question of the assessments. The study used two analysts in coding the language of the questions on each test and then compared their categorizations to increase inter-rater reliability (Merriam, 2009). To increase the coders' reliability, a double-rater read behind consensus model was used based on the work of Miles, Huberman, and Saldana (2014). Both analysts were trained utilizing the Webb training manual (2005) on how to code each test question. The analysts used Mayring's Step Model with protocols to guide their work. Reliability was ensured through the utilization of the double-rater read behind consensus model.

Data Collection

Qualitative Data

The qualitative data collected in this study consisted of publicly available question prompts from the aforementioned TerraNova and IOWA English/Language Arts Practice Tests for grade 8. The prompts were located on two websites accessible through and Internet search using the following search terms: TerraNova and IOWA Practice Tests. The websites were 1) http://classroom.jc-schools.net/8th/, and 2) https://www.ixl.com/standards/iowa/ela/grade-8. In total, the data included two-hundred twenty (220) questions from grade 8 English Language Arts (ELA) practice tests. The breakout of questions contained in each of the two standardized practice assessment instruments is as follows:

- Number of TerraNova English/Language Arts Practice Test Questions: 75
- Number of IOWA English/Language Arts Practice Test Questions: 145

The developers of the IOWA Standardized Assessment (published by Houghton Mifflin Harcourt) claim the test contains measures aligned to broad content standards and is reflective of a level of cognitive complexity appropriate for that child's stage of development. The developmental appropriateness is guided by research and practice in the achievement domains – the major domains of the Common Core State Standards – English Language Arts and Mathematics (Welch & Dunbar, 2014). Tienken and Orlich (2013) stress the importance of setting standards and achievement targets based on cognitive development. The publishers of both the TerraNova and IOWA tests claim the assessment instruments are aligned to the CCSS.

Both the TerraNova and IOWA tests pose questions which are aligned – as claimed – to the CCSS. The tests contain a variety of examination questions requiring student responses such as matching, multiple choice, short answer, essay, and document-based questions (DBQ), also

known as data-based questions. DBQ can be in essay or short-answer questions that are constructed by calling upon one's own knowledge supported by other sources provided, such as maps, excerpts, tables, and charts. In this study, I examined each type of question and the language contained within using Webb's DOK to determine the level of cognitive complexity and depth of knowledge one would need to summon in response. The sampling of all two-hundred twenty (220) questions in this study provided a sample size equaling one-hundred percent (100%) of the total number of questions.

Quantitative Data

The quantitative data were collected as the qualitative content analyses were being completed (this is the convergent, parallel mixed-method aspect of the study). As each question was coded and categorized, the coders kept a frequency count of how many questions were categorized in each of Webb's levels. The frequency data were converted to raw numbers and percentages for each level of Webb's DOK and for aggregate results within two larger categories titled "lower-level" and "higher-level"

Methods

I conducted a convergent, parallel mixed-methods study with qualitative content analysis methods and quantitative descriptive statistics to (a) describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 compares with the language associated with higher-order thinking found in research literature and (b) to describe and compare the complex thinking requirements found on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8.

Qualitative Methods

I used Webb's Depth of Knowledge framework to categorize the language for each question of the assessments. The study used two analysts in coding the language of the questions on each test and then compared their categorizations to increase inter-rater reliability (Merriam, 2009). To increase the coders' reliability, a double-rater read behind consensus model was used based on the work of Miles, Huberman, and Saldana (2014). Both coders/analysts were trained utilizing the Webb training manual (2005) on how to code each test question. The analysts used Mayring's Step Model with protocols to guide their work. Again, the double-rater read behind consensus model was employed to ensure reliability.

Francis (2017) posited that Webb's DOK is an effective and useful tool for teaching and learning and for higher order thinking. Webb's DOK categorizes the levels of thinking students are expected to demonstrate which include all the taxonomies that categorize higher order thinking, and it marks and measures higher order thinking. Karen Hess (2013) stated that the Depth of Knowledge (DOK) is one of the key tools educators need to employ. The tool assists educators in better analyzing the cognitive expectation demanded by the standards, curricular activities, and assessment tasks. This study is limited to: test items contained in the 8th grade English Language Arts sections of the TerraNova and Iowa Tests Practice Tests; coding using Webb's DOK; the time frame of the research; and the determinations derived from the data collected in the document analysis.

Each question in this study was rated as DOK Level 1, 2, 3, or 4 based on its corresponding DOK level. Mayring's step model served as a guide in devising a coding agenda. DOK definitions, examples, and coding rules were considered as per descriptions in the Webb Alignment Tool (WAT) training manual. Two coders coded each question on its depth of

knowledge complexity as per Webb's coding protocol (Sato, Lagunoff, & Worth, 2011; Yuan & Le, 2012).

I used Webb's DOK for category application with regard to the questions found on the aforementioned TerraNova and IOWA English/Language Arts Practice Tests for grade 8. Mayring's Step Model (Mayring, 2000) was utilized in order to ensure deductive category application was followed. Categorization of thinking as prescribed in Webb's DOK represented the pre-existing categories for this study. The following figure describes the step model of deductive category application as per Mayring (2000). This model was adapted for this study to describe this study's coding and analysis process. As a first step, a coding agenda was developed, a protocol established, and definitions noted based on Webb's Depth of Knowledge (DOK) – "a process and criteria for systematically analyzing the alignment between standards and standardized assessments" (Mississippi State University, 2009, p. 5). Subsequently, two expert coders and I took part in a consultant coder training virtual conference call related to coding agenda, rules, protocol, practice coding, and calibration as per Webb's DOK criteria. The expert coder and I – once trained – conducted qualitative content/mixed-methods/document analysis of the 8th grade TerraNova and IOWA standardized practice test questions in English/Language Arts for grade 8 using deductive category application based on Webb's DOK. Next, a final coding and consensus meeting was held by the coders. In conclusion, data analysis and interpretation ensued.

Quantitative Methods

As each question was coded and categorized, the coders created a frequency chart to count the number of questions categorized in each of Webb's levels. The frequency data were converted to raw numbers and percentages for each level of Webb's DOK and for aggregate

results within two larger categories titled “lower-level” and “higher-level” by creating fractions that represented the number of questions in each DOK level and “lower-level” and “higher-level” categories in the numerator and the total number of questions in the denominator.

Mayring’s Step Model:

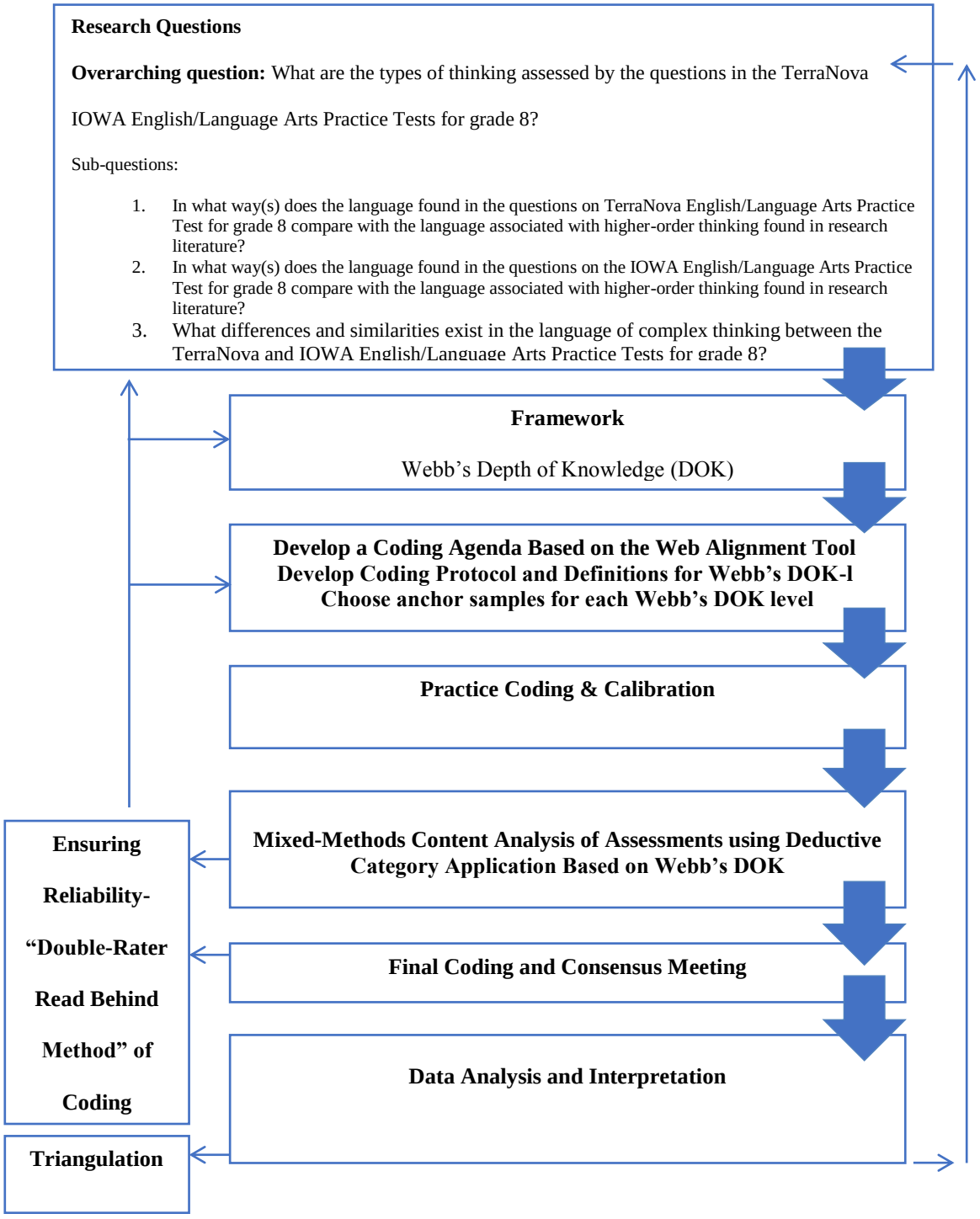

Figure 1. Step model for deductive category application, adapted from Mayring (2000)

Consultant Coder

This study utilized a two-coder system. I worked with one expert coder on the English/Language Arts questions. The expert coder with whom I worked has earned an Ed. D. in Education Leadership, Management, and Policy and her skilled assistance in coding has been summoned in other studies. Her professional and educational background as well as her expertise in school leadership, teaching and instructional practices, and curriculum and assessment served as valuable assets in the completion of this study's coding process.

Coding Scheme

I used Webb's (1997; 2007) four DOK levels as in the Coding Agenda for English Language Arts (ELA) Standards as the guide to coding the questions on the TerraNova and IOWA English Language Arts Practice Tests for grade 8. Coding was based on Webb's DOK Levels and explanations as per Table 2.

Table 2

Webb's DOK Levels with Explanation

DOK Level	Title of Level/Explanation
	Recall and Reproduction
1	Recall elements and details of story structure, such as sequence of events, character, plot, and setting. Items at this level require a student to recall a simple definition, term, fact, procedure, or algorithm.

Skills and Concepts

2 Identify cause and effect, and understand the main idea and purpose implied by text.

Items at this level require a student to develop some mental connections and make decisions on how to set up or approach a problem or activity to produce a response.

Strategic Thinking

3 Evaluate the relative accuracy and usefulness of information from different sources.

Items at this level require a student to engage in planning, reasoning, constructing arguments, making conjectures, and/or providing evidence when producing a response. Items at this level require some complex reasoning and connections to be made.

Extended Thinking

4 Locate, gather, analyze and evaluate written information for the purpose of drafting a reasoned report that supports and appropriately illustrates references and conclusions drawn from research.

Items at this level require a student to engage in complex planning, reasoning, conjecturing, and development of lines of argumentation. Items at this level require a student to make multiple connections between several different key

and complex concepts.

Depth of knowledge considers a number of forms of knowledge such as knowledge based on facts – declarative knowledge, and practical know-how – procedural knowledge. These two forms of knowledge are linked and they form the base that structures creative and strategic thinking opportunities. Levels 1 and 2 of Webb's DOK center on declarative and procedural knowledge (recall, reconstruction). Levels 3 and 4 of Webb's DOK include creative and strategic thinking. DOK levels 3 and 4 afford students opportunities for deeper, analytical, and divergent types of thinking (Sforza, Tienken & Kim, 2016). Researchers have posited that creativity and strategic higher-order thinking occur at DOK Levels 3 and 4.

Webb's Depth of Knowledge (DOK) provides a respected framework with correlation to cognitive complexity. Sforza, Tienken, and Kim (2016) deem Webb's DOK to be a way to define and categorize cognitive complexity of curriculum standards and tasks. The DOK is employed in a number of states and school districts to determine the depth of standardized assessment and state/curriculum standards vis-a-vis their correlation to higher-order thinking/cognitive complexity. In that Webb's DOK is utilized in this study to categorize how the language found in the questions on the TerraNova IOWA English/Language Arts Practice Tests for grade 8 compare to the language associated with higher-order thinking found in the research literature, the study's conclusions can be important in assisting teachers plan and deliver curriculum/instruction and assessment that promotes a greater range of cognitive demand upon students.

Karen Hess (2013) noted that Norman Webb's Depth-of-Knowledge (DOK) framework is an important tool educators can use to analyze the higher-order thinking and cognitive complexity required by classroom activities, content standards, and assessments. Hess (2013) further articulated

the model with content specific descriptions for use by classroom teachers and organizations conducting alignment studies.

Norman Webb's alignment system (Webb, 1997) is one of several alignment systems developed in the relatively recent past. In the Webb alignment system, five criteria for alignment between standards and assessments are described. Webb's study on Depth of Knowledge Consistency and Balance of Representation generated the following criteria:

- Categorical Concurrence: This criterion measures the extent to which the same or consistent categories of content appear in the standards and the assessments. The criterion is met for a given standard if there are more than five assessment items targeting that standard.
- Depth-of-Knowledge Consistency: This criterion measures the degree to which the knowledge elicited from students on the assessment is as complex within the context area as what students are expected to know and do as stated in the standards. The criterion is met if more than half of targeted objectives are hit by items of the appropriate complexity.
- Range-of-Knowledge Correspondence: This criterion determines whether the span of knowledge expected of students on the basis of a standard corresponds to the span of knowledge that students need in order to correctly answer the corresponding assessment items/activities. The criterion is met for a given standard if more than half of the objectives that fall under that standard are targeted by assessment items.
- Balance of Representation: This criterion measures whether objectives that fall under a specific standard are given relatively equal emphasis on the assessment.
- Source of Challenge: This criterion is met if the primary difficulty of the assessment items is significantly related to students' knowledge and skills in the content area as represented in the standards (Webb et al., 2005).

Coding Process, Data Analysis, and Reliability

I used Webb's Depth of Knowledge framework to categorize the language for each of the 220 questions contained in the assessments analyzed in this study. The study used two analysts/coders in coding the language of the questions on each test and then compared their categorizations to increase inter-rater reliability (Merriam, 2009). To increase the coders' reliability, a double-rater read behind consensus model was used based on the work of Miles, Huberman, and Saldana (2014). Utilizing the Webb Training Manual (2005), both analysts/coders were trained on how to code each test question in an hour and a half training session conducted by Dr. Christopher Tienken on March 23, 2020. The analysts used Mayring's Step Model with protocols to guide their work. Reliability was established via the double-rater read behind consensus model. The coding data of both coders was then tallied into a data collection chart as exemplified in Table 3.

Table 3

DOK Data Collection Chart

TerraNova ELA	No. of Items	DOK Level 1	DOK Level 2	DOK Level 3	DOK Level 4	% Levels 1 & 2	% Levels 3 & 4
Grade 8	75						
IOWA ELA							
Grade 8	145						

Once data were entered for all items, answers to research questions/sub questions one and two were derived. Following this, data were gleaned to answer the third research question as exemplified by Table 4.

Table 4

Comparison of the TerraNova and IOWA English/Language Arts Practice Tests

DOK Levels	% Levels 1 & 2	% Levels 3 & 4
TerraNova Grade 8		
IOWA Grade 8		

On March 23, 2020, the expert coder and I participated in a virtual conference/training-coding session with Dr. Christopher H. Tienken – Associate Professor, Department of Education Leadership, Management, and Policy, Seton Hall University. Discussion of each level of Webb's DOK shed light as to the types of questions that would be coded to a corresponding level of the DOK. The coding process was enhanced by this discussion as coders were now able to recognize key words/phrases found in the questions and relate them to a corresponding level of Webb's DOK. As the training progressed, 35 of the 145 questions on the IOWA English/Language Arts Practice Test were analyzed and coded with the trainer monitoring the expert coder and me. This assisted the coders to become more familiar with and exacting in the coding process. In the training, coders became familiar with the progression of cognitive rigor – from lower level to higher level – as per Webb's DOK. While working with the trainer, the expert coder and I attained 90% agreement on the categorization scores. The discussions about each analyzed question under the guidance of the trainer helped to ensure that coders analyzed and coded questions objectively. Consensus was gained and validity and reliability were ensured via the inputted explanation for each of the four levels of Webb's DOK as well as by employing the double read consensus model and inter-rater reliability. As defined by Lange (2011), inter-rater reliability is the extent to which two or more raters (or observers, coders, examiners) agree. It addresses the issue of consistency of the implementation of a rating system. Inter-rater reliability can be evaluated by using a number of different statistics.

During the initial conference call/virtual coder training session, the expert coder and I alone reviewed the Webb Alignment Tool and commenced calibrating a set of 10 questions on the IOWA ELA Practice Test to levels depicted in Webb's DOK. The first order of business in the session was to view the DOK level by level with adherence to Webb's (2005) Web Alignment Tool. Review of the DOK levels enlightened the coders with the overlapping of the DOK with Bloom's Taxonomy and Hess's Cognitive Rigor Matrix. Discussion between the expert coder and I initiated a consensus relative to coding protocol.

After the training session on March 23, 2020, the expert coder and I resumed coding an additional 20 questions (2 sets of 10) found in the English/Language Arts section of the IOWA Practice Test for grade 8. Webb's DOK levels served as the rubric for categorizing the questions by code. Questions deemed by agreement as Level 1 were graded 1; Level 2 questions were graded 2, etc. As coding transpired, the coders discussed the scoring of each question as per its score on the DOK. After analyzing and discussing each set of 10 questions, the coders reached one-hundred percent (100%) agreement using the double-rater read-behind consensus method and memorialized in a coding table developed by Sydoruk (2018) and modified by me to befit this study. The questions whereby one coder scored a question as Level 1, and the other as Level 2, the score of Level 2 was given; by doing this, the two coders organized each question into its DOK level. Table 5 is an example of the final tally coding table adapted for this study:

Table 5

Final Scoring Chart

Terra Nova	# of Items	DOK Level 1	DOK Level 2	DOK Level 3	DOK Level 4	% Level 1	% Level 2	% Level 3	% Level 4
ELA									
IOWA	# of Items	DOK Level 1	DOK Level 2	DOK Level 3	DOK Level 4	% Level 1	% Level 2	% Level 3	% Level 4
ELA									

Coders agreed that questions scoring in the DOK levels 3 and 4 would be deemed higher level as per the Web Alignment Tool (Webb, Alt, Eli, & Vesperman, 2005). Questions scoring in the 3 to 4 range call upon students to use higher-order thinking due to the cognitive complexity commensurate with questions/problems as categorized in this range of the DOK. Questions categorized as 1's and 2's on the DOK were deemed lower level as cognitive complexity/rigor was not evident, and they did not require the test taker to employ higher-order thinking.

During the same coding session, the expert coder and I conducted analysis of 20 more questions on the IOWA ELA Practice Test. Each of the 20 questions, reviewed in sets of 10, were categorized in line-by-line coding fashion and individually scored in accordance with the language of Webb's DOK levels. Another step we took to increase the coders' reliability was a double-rater read behind consensus model which proved effective in coding standards for other studies (Miles, Huberman, & Saldaña, 2014; Sato, Lagunoff, & Worth, 2011; from Sforza, Tienken, & Kim, 2016).

To guarantee coding reliability, the expert coder and I adhered to the double-rater read behind consensus model as we compared categorized DOK scores for each question reviewed. Before the initial training session, I practiced calibration following the double-rater read behind consensus model

with the guidance of an experienced independent coder. In the calibration meetings, deductive coding drills were practiced by the coders.

The expert coder and I convened via a Zoom virtual session for a second calibration session on March 25, 2020. The two expert coders and I met to calibrate scores for 35 questions. Discussion about each question, and the rationale for assigning particular DOK scores to the questions was conducted. By way of established rapport and communication, consensus was established relative to the DOK scores each coder recorded. This was accomplished via comments entered in a data collection Google Doc, email exchanges, virtual person-to-person Zoom sessions, and phone conferencing. Inter-rater reliability was increased by utilization of the double-rater read behind consensus model which – in turn – enhanced inter-rater reliability and afforded a way to calculate data and monitor agreement amongst coders (Miles, Huberman, & Saldaña, 2014; Sato, Lagunoff, & Worth, 2011; from Sforza, Tienken, & Kim, 2016).

Working on sets of 10 questions, the two expert coders and I attained ninety percent (90%) consensus in coding scores. Then the expert coder and I attained agreement on each set of ten questions in the coding session subsequent to employing the double-rater read behind strategy and referencing Webb's Training Manual. Each coder justified her/his categorization of each question's score in order to come to an agreement. The percentage of agreement was calculated and recorded after the coding of each set of questions reviewed. We highlighted questions that were scored differently in order to revisit them for further discussion and ultimate agreement. As per Vanderhook (2020), a protocol for deliberations was followed:

- The team members explained their respective categorization scores.
- A review of Webb's DOK language was compared to language contained in the question under deliberation.

- Exemplar questions related to DOK levels were reviewed by the team.
- As per Webb's Web Alignment Tool (WAT), a higher rating/score was recorded for questions that did not reach team consensus, e.g. if one coder scored a particular question/problem as a Level 1 (recall) per Webb's DOK, and the other coder scored it as a 2 (skill/concept), the question would be scored as a 2 (skill/concept).

This study's protocol was modeled in line with like studies to provide topic area methodological consistency (Miles, Huberman, & Saldaña; Sato, Lagunoff, & Worth, 2011; from Sforza, Tienken, & Kim, 2016). Then, team members reviewed and coded the balance of questions individually and compared results at a later time.

Webb's DOK was used as a reliable way in which to code the TerraNova and IOWA English/Language Arts Practice Tests and determine scores of each question/problem. The Webb Alignment Tool (Webb, Alt, Ely, & Vesperman, 2005), and the DOK for English/Language Arts gave the coders detail as to the cognitive complexity – or lack thereof – of the questions/problems reviewed and coded. Also, descriptions of the types of questions provided insight as to the levels of the questions/problems reviewed. As per Webb, et al (2005), the following describes the types of thinking one would employ in response questions as viewed in line with a DOK score:

- English/Language Arts - DOK Level 1 (Recall): Remember elements and details of story structure, such as sequence of events, character, plot and setting; represent in words or diagrams a concept or relationship; use punctuation marks correctly; describe the features of a place or people.
- English/Language Arts - DOK Level 2 (Skills/Concepts): 1dentify and summarize the major events in a narrative; use context cues to identify the meaning of unfamiliar words; describe

the cause/effect of a particular event; use contextual clues to identify unfamiliar words; identify and summarize main points.

- English/Language Arts - DOK Level 3 (Strategic Thinking): Support ideas with details and examples; use voice appropriate to the purpose and audience; determine the author's purpose and describe how it affects the interpretation of a reading selection; critically analyze literature; edit writing to produce logical progression.
- English/Language Arts - DOK Level 4 (Extended Thinking): Analyze and synthesize information from multiple sources. Describe and illustrate how common themes are found across texts from different cultures; analyze and synthesize information from multiple sources; create compositions that synthesize, analyze, and evaluate; examine and explain alternative perspectives across sources (Webb, et al, 2005).

On March 26, 2020, a third calibration meeting was convened. In this session, the expert coder and I held coding session with discussions about the remaining 60 questions of the IOWA ELA Practice Test in sets of 20. Reliability of the coding was ensured via the utilization of the double-rater read behind consensus model. Coders attained one-hundred percent (100%) agreement in each set of 20 questions using the double rater read behind consensus model. Subsequent to discussion, questions requiring changes in score were agreed upon. For example:

Read the passage. Then, select the synonym of the word in bold.

Within the company, Mrs. Spencer was known for being a ***prudent*** *leader who was able to assess situations and understand potential outcomes. She always made cautious decisions that balanced short - and long-term goals while avoiding risk.* (https://www.ixl.com/ela/grade-8/determine-the-meaning-of-words-using-synonyms-in-context)

The coders had initial disagreement concerning the question above (taken from the IOWA ELA Practice Test) as one coder categorized the question as a Level 2, the other coder as a Level 1. Using the double-rater read behind consensus protocol, the question was ultimately scored as a Level 2.

On March 27, 2020, a fourth virtual/Zoom conference scoring/calibration session was held between the expert scorer and me. At this session, the focus was on analyzing questions contained in the English/Language Arts (ELA) section of the TerraNova Practice Test. In the TerraNova ELA Practice Test section, the coders completed 38 questions with a 100% total agreement in which several questions had DOK levels lowered and several elevated. On March 28, 2020, the coders completed scoring/calibrating the remaining 37 questions of the TerraNova ELA Practice Test. Some DOK Level scores were lowered and some elevated using the consensus model. For example:

Choose the best order for the sentences.

1. Exhausted from the grueling doubleheader, the Bluejays team members sat on the bench to listen to the coach. 2. After winning the first game, the Bluejays celebrated with oranges and bagels, resting for the next game. 3. Anxious teammates arrived at the field at 7:30 a.m. for an early morning practice prior to their games. 4. The Bluejays struggled to beat the Cardinals in the first game of the doubleheader. 5. The Bluejays miraculously went on to win the first game by a narrow margin. 6. Regrettably, in the second game the Bluejays lost to the Indians by four runs. 7. In the third inning, the Cardinals scored three runs when their best hitter slammed one out of the park.

A. 3 - 4 - 7 - 5 - 6 - 2 - 1

B. 3 - 4 - 7 - 5 - 2 - 6 - 1

C. 4 - 3 - 7 - 5 - 2 - 6 - 1

D. 4 - 3 - 5 - 7 - 2 - 6 - 1 (http://classroom.jc-schools.net/8th/8th-LA.pdf)

While analyzing the question above, the coders were in disagreement – one coder scoring it as Level 2, the other as Level 3. Subsequent to discussion, and following the double-rater read behind consensus model, the coders agreed that the question is a Level 3.

During the virtual scoring session of March 28, 2020 – after completing scoring of the TerraNova ELA questions – the coders conducted a final review of particular questions from both practice tests that had been discussed in prior coding sessions. The coders were is agreement on 100% of the scores that had been given for all discussed questions from both practice tests. An agreement was made that – in the event questions arise as the study moves forth – the coders would reconvene to calibrate and re-score.

Chapter Summary

Chapter III has provided information as to the coding process and protocol used in this study which investigated how the language found in the questions on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8 compare with the language associated with higher-order thinking found in research literature. Webb's DOK was used as the model for coding the questions in order to ascertain the level of cognitive complexity the questions/problems offered. The research questions were answered via a mixed-method content analysis research methodology which produced quantitative and quantitative data. The research process was guided using the step model (Mayring, 2000) for deductive category application. The step model is a visual depiction of this study's process/protocol and methods employed to ensure the study's credibility. In this chapter, coding rules, definitions related to the methodology utilized in this study were presented as they gave guidance in assuring an organized coding agenda.

With an analytical focus on answering my research questions, I will present the findings of this study in Chapter IV.

Chapter IV

Results

Introduction

Chapter IV presents the findings of the convergent, parallel mixed-methods study to describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts (ELA) Practice Tests for grade 8 compared with the language associated with higher-order thinking found in the research literature. A sample size of 220 questions was analyzed in this study. Two trained coders held six (6) coding sessions and utilized the double-rater read-behind consensus model to discuss and categorize each question contained in the tests. Two trained coders held six (6) coding sessions. During these coding sessions, Webb's Depth of Knowledge (DOK) schema was used as an alignment tool in which to measure the cognitive complexity/higher-order thinking one would need to engage in answering the various ELA questions.

The coders agreed in their first calibration session that questions placed into the third and fourth levels of Webb's Depth of Knowledge, as identified in Webb's DOK would be considered higher order thinking. Higher-order thinking categories included Levels 3 (Strategic Thinking) and 4 (Extended Thinking) of Webb's DOK. Questions/problems that were categorized as Levels 1 (Recall) and 2 (Skills and Concepts) as per Webb's DOK were not considered as the types of questions/problems that required higher-order thinking skills necessary for one to respond successfully. The coders utilized the double-rater read-behind consensus model to reach alignment of each of the questions provided by the aforementioned TerraNova and IOWA Practice Tests reviewed in this study. The double-rater read behind consensus model allowed for the coders to discuss the placement of each question along Webb's DOK, and it assisted in the

development of a rationale for categorization. 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). During each calibration session, the coders discussed the placement of each question into a specific category and examined differences in placement until a consensus was reached. After coding and discussing each question, the total number of questions in each category of Webb's DOK and the percentages were calculated.

The research design for this study was a convergent, parallel mixed-methods study with qualitative content analysis methods and quantitative descriptive statistics. "The purpose of content analysis is to organize and elicit meaning from the data collected and draw realistic conclusions from it" (Bengtsson, 2016, p. 8). "As a research method, it represents a systematic and objective means of describing and quantifying phenomena. A prerequisite for successful content analysis is that data can be reduced to concepts that describe the research phenomenon by creating categories, concepts, a model, conceptual system, or conceptual map" (Elo et al., 2014, p. 1). Creswell and Plano-Clark (2011) described a convergent parallel design as one whereby the researcher simultaneously conducts the quantitative and qualitative elements in the same phase of the research process, weighs the methods equally, analyzes the two components independently, and interprets the results together. Figure 2 depicts the convergent, parallel mixed methods design:

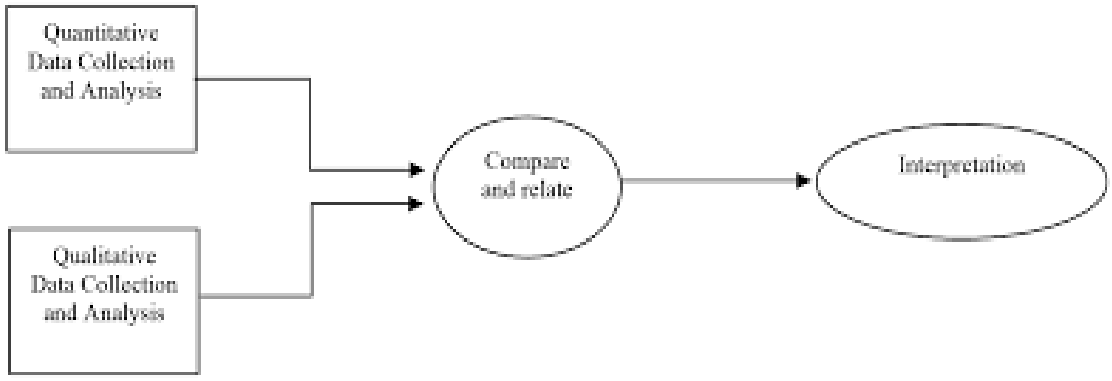

Figure 2. Image depicting convergent, parallel mixed methods design.

This study was guided by the overarching question: What are the types of thinking assessed by the questions in the TerraNova and IOWA English/Language Arts Practice Tests for grade 8? There were three sub-questions associated with the overarching question. These sub-questions were the key questions used for interpreting the data derived in this study.

Findings for Sub-question 1

The first sub-question posed in this study was: In what way(s) does the language found in the questions on the TerraNova English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in the research literature? In accordance with Webb's DOK, the lowest level of cognitive complexity/higher-order thinking is Level 1 - recall. In this study questions scored and categorized as DOK Levels 1 and 2 were both deemed not to contain the type of language that requires one to be challenged with cognitive complexity nor for one to engage in higher-order thinking. Questions/problems scored and categorized as Levels 3 and 4 as per Webb's DOK were considered to contain the type of language that challenges one with cognitive complexity and necessitates that one engages in higher-order thinking. Although both DOK Levels 3 and 4 questions and problems were considered as containing cognitive complexity, the coding team agreed that Level 4 would be considered the highest level vis-a-vis higher-order thinking.

Qualitative Findings for Sub-question 1 –TerraNova Grade 8 English/Language Arts Practice Test

The expert coder and I analyzed Seventy-five (75) TerraNova English/Language Arts questions. The 75 questions were arranged in the practice test as follows: Language Arts Test – 36 questions; Reading Test – 39 questions. The coders utilized these two sections to encompass

the broad heading of English/Language Arts. Of the 75 questions, the following scores as per Webb's DOK were determined. There were seven (7) questions that found the coders in disagreement. Using the double-rater read behind consensus model, ultimately 100% agreement was attained. Of the seven questions of disagreement, four questions initiated discussion as to whether or not a score of DOK 1 or DOK 2 was merited. Three of the questions initiated discussion as to whether or not a score of DOK 2 or DOK 3 was merited.

Examples of these questions requiring discussion as per the double-rater read behind consensus model are as follows:

TerraNova Question #40

Read the passage and complete Blank 2.

Many people feel that it is cruel to keep animals in the zoo. Those who feel this way would explain that animals kept in __1__ have a poor quality of life because they don't have the freedom that they would experience in the wild. However, supporters of zoos argue that the only way many endangered animals will replenish their numbers is with the __2__ that they receive because they are in a safe environment. Fortunately, many zoos and wildlife parks are striving to create authentic living environments for their animals. One can see that both sides of the argument have relevancy.

A. challenge

B. threat

C. donations

D. protection

(http://classroom.jc-schools.net/8th/8th-read.pdf)

The coders had initial disagreement concerning the question above (taken from the TerraNova ELA Practice Test) as one coder categorized the question as a Level 2, the other coder as a Level 1. Using the double-rater read behind consensus protocol, the question was ultimately scored as a Level 2. The coders reviewed exemplary questions from the Webb's Alignment Tool, WAT, re-read the descriptors provided within Webb's DOK and considered what each test was asking from a student's point of view. Also, the WAT requires that when a question or prompt includes two levels of thinking, the coders must award the higher level to the question.

Utilizing Webb's Alignment Tool/Webb's Depth of Knowledge Overview (2005) the coders came to agreement that language found in the question and the thinking challenge in which one would employ fit the following criteria for DOK Level 2:

> Level 2 includes 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. Some important concepts are covered, but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. [Additionally] Items at this level require a student to develop some mental connections and make decisions on how to set up or approach a problem or activity to produce a response (Webb, et al, 2005).

Using Hess' Cognitive Rigor Matrix & Curricular Examples (Hess et al., 2009) that applies Webb's DOK Levels to Bloom's Cognitive Process Dimensions for ELA, to additionally validate the Level 2 score, the following characteristics applied to the question:

- Make basic inferences or logical predictions from data or texts (Bloom's Taxonomy – Understand).
- Use context to identify meaning of words/phrases (Bloom's Taxonomy – Apply)

As previously noted, Hess' Cognitive Rigor Matrix was only used as an additional validation of a DOK score only. Webb's DOK was the sole instrument used to ultimately determine the categorization of a question in the TerraNova and IOWA English/Language Arts Practice Tests analyzed in this study.

TerraNova Question #15

Choose the best order for the sentences:

1. As I sat down, Mr. Spurlin glanced at the clock.

2. While staring into my fearful eyes, Mr. Spurlin pointed toward the door.

3. I walked into the classroom and headed toward my desk.

4. As our eyes met, I knew I was in trouble.

5. Mr. Spurlin turned from the clock to me.

6. The clock read ten minutes after nine.

7. I got up from my desk and walked out the door.

A. 3 - 1 - 6 - 5 - 4 - 2 - 7

B. 3 - 1 - 5 - 6 - 4 - 7 - 2

C. 6 - 3 - 5 - 1 - 4 - 2 - 7

D. 6 - 3 - 1 - 4 - 5 - 2 – 7

The coders disagreed initially on the coding of the question. One coder scored it as Level 2, the other as Level 3. Subsequent to discussion, and following the double-rater read behind consensus model, the coders agreed that the question is a Level 3 because if there is disagreement on the score of a question, the question is scored as the higher of the two coder scores.

Utilizing Webb's Alignment Tool/Webb's Depth of Knowledge Overview (2005) the coders came to agree that language found in the question and the thinking challenge in which one would employ fit the following criteria for DOK Level 3:

> Deep knowledge becomes a greater focus at Level 3. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Standards and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Additionally, items at this level require a student to engage in planning, reasoning, constructing arguments, making conjectures, and/or providing evidence when producing a response. Items at this level require some complex reasoning and connections to be made (Webb, et al, 2005).

Validating the score using Hess' Cognitive Rigor Matrix & Curricular Examples (Hess et al., 2009) that applies Webb's DOK Levels to Bloom's Cognitive Process Dimensions for ELA the question contained the following characteristics:

- Explain, generalize or connect ideas using supporting evidence (Bloom's Taxonomy – Understand)

- Revise a draft for meaning or progression of ideas (Bloom's Taxonomy – Apply)
- Use reasoning and planning (Bloom's Taxonomy – Analyze) (Hess, et al, 2009).

Quantitative Findings for Sub-Question 1 – TerraNova Grade 8 English/Language Arts Practice Test

Quantitative findings for the TerraNova English/Language Arts Practice test analyzed in this study are as follows:

- Number of questions categorized as DOK Level 1 – 53 (71%)
- Number of questions categorized as DOK Level 2 – 19 (25%)
- Number of questions categorized as DOK Level 3 – 3 (4%)
- Number of questions categorized as DOK Level 4 – 0 (0%)

Findings for Sub-question 2

The second sub-question posed in this study was: In what way(s) does the language found in the questions on the IOWA English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in research literature?

Qualitative Findings for Sub-question 2 – IOWA Grade 8 English/Language Arts Practice Test

The coders analyzed 145 IOWA English/Language Arts questions. The 145 questions were arranged in sections as per the skill set tested, e.g. demonstrate command of the conventions of standard English grammar and usage when writing or speaking; recognize and correct inappropriate shifts in verb voice and mood; spell correctly; etc. The entire link utilized by the coders in rating the questions on the IOWA ELA Practice Test was entitled Language Arts. Of the 145 questions, the following scores as per Webb's DOK were determined. There were twelve (12) questions that found the coders in disagreement. Using the double-rater read

behind consensus model, ultimately 100% agreement was attained. Of the twelve questions of disagreement in the IOWA ELA Practice Test, all questions initiated discussion as to whether or not a score of DOK 1 or DOK 2 was merited. Following is an example of a question that included characteristics of DOK 1 and DOK 2:

IOWA Question #24

Complete the sentence with the correct comparative or superlative adjective.

Of all the contestants on the game show, the gentleman from Detroit had

the  broad knowledge of international affairs.

(https://www.ixl.com/ela/grade-8/form-and-use-comparative-and-superlative-adjectives)

The coders had initial disagreement concerning the question above (taken from the IOWA ELA Practice Test) as one coder categorized the question as a Level 2, the other coder as a Level 1. Using the double-rater read behind consensus protocol, the question was ultimately scored as a Level 2 because if there is disagreement on the score of a question, the question is scored as the higher of the two coder scores.

Utilizing Webb's Alignment Tool/Webb's Depth of Knowledge Overview (2005) the coders came to agree that language found in the question and the thinking challenge in which one would employ fit the following criteria for DOK Level 2:

> Level 2 includes 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. Some important concepts are covered, but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A

> Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. Additionally, items at this level require a student to develop some mental connections and make decisions on how to set up or approach a problem or activity to produce a response (Webb, et al, 2005).

In accordance with Hess' Cognitive Rigor Matrix & Curricular Examples (Hess, 2009) that applies Webb's DOK Levels to Bloom's Cognitive Process Dimensions for ELA, the score was additionally validated as it contained the following Level 2 characteristics:

- Make basic inferences or logical predictions from data or texts (Bloom's Taxonomy – Understand).
- Use context to identify meaning of words/phrases (Bloom's Taxonomy – Apply) (Hess, et al, 2009)

Quantitative Findings: IOWA Grade 8 English/Language Arts Practice Test

Quantitative findings for the IOWA English/Language Arts Practice test analyzed in this study are as follows:

- Number of questions categorized as DOK Level 1 – 115 (79%)
- Number of questions categorized as DOK Level 2 – 30 (21%)
- Number of questions categorized as DOK Level 3 – 0 (0%)
- Number of questions categorized as DOK Level 4 – 0 (0%)

Findings for Sub-question 3

The third sub-question posed in this study was: What differences and similarities exist in the language of complex thinking between the TerraNova and IOWA English/Language Arts Practice Tests for grade 8?

Qualitative Findings

The majority of the language in questions from the TerraNova and IOWA ELA Practice Tests align with the language of lower level thinking found in the research literature. Substantial percentages of questions categorized as DOK Level 1 indicate that both tests contain a preponderance of lower-level thinking questions. Examples of DOK Level 1 questions on both practice tests are below.

TerraNova Question #1

Which of the following is written correctly?

A. Kira baked a cake it tasted awful.

B. Kira baked. A cake. It tasted awful.

C. Kira baked a cake. It tasted awful.

D. Kira baked, a cake; it tasted awful. (http://classroom.jc-schools.net/8th/8th-LA.pdf)

From the IOWA ELA Practice Test – question number 10

Which of the following contains a vague pronoun reference?

How are they working to improve technologies that turn the sun's energy directly into electricity?

How are scientists working to improve technologies that turn the sun's energy directly into electricity?

Submit (https://www.ixl.com/ela/grade-8/identify-vague-pronoun-references)

Both questions, as were all questions analyzed that were coded as DOK Level 1 fit criteria outlined in Webb's Alignment Tool and Overview (2005). Level 1 ELA (Reading and Writing) questions, as per Webb's (2005, p. 70) DOK, contain the following recall characteristics:

Reading Level 1: Level 1 requires students to receive or recite facts or to use simple skills or abilities. Oral reading that does not include analysis of the text, as well as basic comprehension of a text, is included. Items require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Support ideas by reference to verbatim or only slightly paraphrased details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage (Webb, et al, 2005, p. 73).

Writing Level 1: Level 1 requires the student to write or recite simple facts. The focus of this writing or recitation is not on complex synthesis or analysis, but on basic ideas. The students are asked to list ideas or words, as in a brainstorming activity, prior to written composition; are engaged in a simple spelling or vocabulary assessment; or are asked to write simple sentences. Students are expected to write, speak, and edit using the conventions of Standard English. This includes using appropriate grammar, punctuation, capitalization, and spelling. Students demonstrate a basic understanding and appropriate use of such reference materials as a dictionary, thesaurus, or Web site. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Use punctuation marks correctly.

- Identify Standard English grammatical structures, including the correct use of verb tenses (Webb, et al, 2005, p. 74).

The majority of the language in questions from the TerraNova and IOWA ELA Practice Tests align with the language of lower level thinking found in the research literature. Significant percentages of questions categorized as DOK Level 2 indicate that both tests contain Level 2/lower-level thinking questions as the second most in number/percentage on the ELA Practice Tests. Examples of DOK Level 2 questions on both practice tests are:

TerraNova Question #5 (Reading Section)

Read the passage and complete Blank 2: Twenty years ago, if you had a grade point average of 3.0 or better, you were almost guaranteed to be admitted to a state college. Today, things are very different. Most high school students would agree that college admissions policies have become much harder. 1. One reason that this has occurred is because a higher percentage of students now attend college immediately after completing high school. Therefore, more students are 2 for the same number of openings. As a result, colleges are now looking at standardized test scores, entrance essays, and students' extra-curricular activities.

A. hoping

B. challenging

C. competing

D. testing (http://classroom.jc-schools.net/8th/8th-read.pdf)

IOWA Question #10

Read the sentences.

The star HD 162826 has been called a solar sibling to our sun. HD 162826 was likely born in the same gas cloud.

Now combine the sentences by completing the relative clause.

The star HD 162826, **which**  was likely born in the same gas cloud.

Submit (https://www.ixl.com/ela/grade-8/combine-sentences-using-relative-clauses)

Both questions, as were all questions analyzed that were coded as DOK Level 2 fit criteria outlined in Webb's Alignment Tool and Overview (2005). Level 1 ELA (Reading and Writing) questions, as per Webb's (2005, p. 71/72) DOK, contain the following skills and concepts characteristics:

Reading Level 2: Level 2 includes 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. Some important concepts are covered, but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a reading selection.

- Identify and summarize the major events in a narrative (Webb, et al, 2005, p. 73).

*Writing Level 2***:** Level 2 requires some mental processing. At this level, students are engaged in first-draft writing or brief extemporaneous speaking for a limited number of purposes and audiences. Students are expected to begin connecting ideas, using a simple organizational structure. For example, students may be engaged in note-taking, outlining, or simple summaries. Text may be limited to one paragraph. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Construct or edit compound or complex sentences, with attention to correct use of phrases and clauses.
- Use simple organizational strategies to structure written work.
- Write summaries that contain the main idea of the reading selection and pertinent details (Webb, et al, 2005, p. 74).

The findings presented above for both the TerraNova and the IOWA ELA Practice Tests for grade 8 indicate a low percentage of DOK Level 3. A low percentage of Level 3 language in the test questions signifies that both the TerraNova and the IOWA ELA Practice Tests for grade 8 contained a range of limited to no DOK Level 3 questions. As noted earlier in this chapter, questions/problems scored and categorized as Levels 3 and 4 as per Webb's DOK were considered to contain the type of language that challenges one with cognitive complexity and necessitates that one engages in higher-order thinking. Of the 75 TerraNova ELA Practice Test questions analyzed and coded, only 3 were scored as DOK Level 3. None of the 145 IOWA ELA Practice Test questions analyzed and coded were scored as DOK Level 3. An example of a DOK Level 3 questions on the TerraNova ELA Practice Test for grade 8 follows:

TerraNova Question #14 (Reading Section)

Choose the best order for the sentences:

1. Early this morning, my father and I decided to go to a concert in the park.

2. On the way home, my father and I agreed that it was the best day we'd ever spent together.

3. After our walk, we headed over to the outdoor concert arena.

4. Once at the park, we took a scenic walk through lush gardens and several art museums.

5. We found two shady seats near the front of the arena.

6. We drove forty miles from our house to the park.

7. As the music began, I closed my eyes and remembered the beautiful things I had seen on our walk.

A. 1 - 6 - 3 - 4 - 5 - 7 - 2

B. 1 - 6 - 4 - 3 - 5 - 7 - 2

C. 1 - 6 - 4 - 3 - 5 - 2 - 7

D. 1 - 6 - 4 - 3 - 2 - 5 – 7 (http://classroom.jc-schools.net/8th/8th-LA.pdf)

The above question analyzed and coded as DOK Level 3, fit criteria outlined in Webb's Alignment Tool and Overview (2005). Level 3 ELA (Reading and Writing) questions, as per Webb's (2005) DOK, contain the following strategic thinking characteristics:

Reading Level 3: Deep knowledge becomes a greater focus at Level 3. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Standards

and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Explain or recognize how the author's purpose affects the interpretation of a reading selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature (Webb, et al, 2005, p. 73/74).

*Writing Level 3***:** Level 3 requires some higher-level mental processing. Students are engaged in developing compositions that include multiple paragraphs. These compositions may include complex sentence structure and may demonstrate some synthesis and analysis. Students show awareness of their audience and purpose through focus, organization, and the use of appropriate compositional elements. The use of appropriate compositional elements includes such things as addressing chronological order in a narrative, or including supporting facts and details in an informational report. At this stage, students are engaged in editing and revising to improve the quality of the composition (Webb, 2005). Some examples that represent, but do not constitute all of, Level 3 performance are:

- Support ideas with details and examples.
- Use voice appropriate to the purpose and audience.
- Edit writing to produce a logical progression of ideas (Webb, et al, 2005, p. 75).

Quantitative Findings

Table 6 presents numbers of questions categorized in the DOK Levels and percentages of occurrence for both the TerraNova and the IOWA ELA Practice Test for grade 8. In general, the table shows the preponderance of lower level thinking questions – DOK levels 1 and 2 – and the scarcity of higher-order thinking questions – DOK levels 3 and 4.

Table 6

Final Scoring Chart

Terra Nova	# of Items	DOK Level 1	DOK Level 2	DOK Level 3	DOK Level 4	% Level 1	% Level 2	% Level 3	% Level 4
ELA	75	53	19	3	0	71	25	4	0
IOWA	# of Items	DOK Level 1	DOK Level 2	DOK Level 3	DOK Level 4	% Level 1	% Level 2	% Level 3	% Level 4
ELA	145	115	30	0	0	79	21	0	0

A comparison of total percentages of lower-level and higher-level thinking questions contained in the TerraNova and the IOWA English/Language Arts Practice Tests are shown in Table 7.

Table 7

Comparison of the TerraNova and IOWA English/Language Arts Practice Tests

DOK Levels	% Levels 1 & 2 Lower-Level	% Levels 3 & 4 Higher-Level
TerraNova Grade 8	96	4

IOWA Grade 8	100	0

Conclusion

My purpose for this convergent, parallel mixed-methods study was to describe the way(s) in which the language found in the questions on TerraNova and IOWA Practice Tests in English/Language Arts for grade 8 compare with the language that promotes higher-order thinking found in research literature. I used Webb's Depth of Knowledge framework to categorize the language for each question of the assessments. Questions scored and categorized as DOK Levels 1 and 2 were both deemed not to contain the type of language that requires one to be challenged with cognitive complexity nor for one to engage in higher-order thinking. Questions/problems scored and categorized as Levels 3 and 4 as per Webb's DOK were considered to contain the type of language that challenges one with cognitive complexity and necessitates that one engages in higher-order thinking.

Overall findings as per all questions analyzed and coded from both practice tests are as follows:

- Of the 220 total questions analyzed and coded, 168 questions were categorized as DOK Level I, equating to 76% of all questions analyzed and scored in this study.
- Of the 220 total questions analyzed and coded, 49 questions were categorized as DOK Level 2, equating to 22% of all questions analyzed and scored in this study.
- Of the 220 total questions analyzed and coded, 3 questions were categorized as DOK Level 3, equating to 2% of all questions analyzed and scored in this study.
- Of the 220 total questions analyzed and coded, no questions were categorized as DOK Level 4, equating to 0% of all questions analyzed and scored in this study.

Chapter IV presented results related to the convergent, parallel mixed-methods study to describe the way(s) in which the language found in the questions on the TerraNova and IOWA English/Language Arts (ELA) Practice Tests for grade 8 compared with the language associated with higher-order thinking found in the research literature. The preponderance of lower-level thinking questions found in the analysis of questions reviewed in this study, results presented in Chapter IV confirms the hypothesis presented in Chapter III.

Chapter V presents a summary of the methodology and discussion of findings as they relate to my research subquestions. Implications vis-à-vis policy, practice, and future research recommendations will also be addressed in Chapter V.

Chapter V

Conclusions

Chapter V provides conclusions, implications for policy and practice, and recommendations for future research. My purpose for this convergent, parallel mixed-methods study was to describe the way(s) in which the language found in the questions on TerraNova and IOWA Practice Tests in English/Language Arts for grade 8 compare with the language that promotes higher-order thinking found in research literature. Creswell and Plano-Clark (2011) described a convergent parallel design as one whereby the researcher simultaneously conducts the quantitative and qualitative elements in the same phase of the research process, weighs the methods equally, analyzes the two components independently, and interprets the results together. In this study, 220 questions were reviewed and analyzed using Webb's DOK (1997). The questions analyzed were found in the TerraNova (http://classroom.jc-schools.net/8th/) and the IOWA (https://www.ixl.com/standards/iowa/ela/grade-8) English/Language Arts Practice Tests for grade 8. This study found no empirical evidence that suggests either of the practice tests examined and analyzed promoted higher-order thinking skills.

Both the TerraNova and the IOWA standardized tests are used widely in private and parochial schools. Some public schools use the tests to assess student achievement in grades K through 2. The TerraNova is also used in some Department of Defense Dependents Schools and – as part of the California Achievement Test (CAT/6) – the state of California utilizes the TerraNova standardized test. Only the state of Iowa uses the IOWA standardized test for federal compliance. Data derived from the TerraNova and IOWA tests are used in private and parochial schools, and in some public schools, to determine students at risk; placement in gifted and talented programs; and high school acceptance.

Methodology Summary

The theoretical framework used in this study is Webb's Depth of Knowledge (DOK). Webb's DOK is a hierarchy demarcated with four levels of cognitive complexity. The Wisconsin Center for Education Products and Services, Inc. (2017) notes thinking requirements as per the Webb's DOK as follows:

- Level 1 – Recall and Reproduction (ELA): Requires only surface understanding of text, often verbatim recall
- Level 2 – Skills and Concepts (ELA): Requires processing beyond recall and observation; requires both comprehension and subsequent processing of text or portions of text; involves ordering, classifying text as well as identifying patterns, relationships, and main points
- Level 3 – Strategic Thinking (ELA): Requires students to go beyond text; requires students to explain, generalize, and connect ideas; involves deep inferencing, prediction, elaboration, and summary; requires students to support positions using prior knowledge and evidence and manipulate themes across passages
- Level 4 – Extended Thinking (ELA): Requires complexity at least at the level of DOK 3, but also an extended time to complete the task, such as conducting a research project over many weeks; a project that requires extended time, but repetitive or lower-DOK tasks are not at Level 4; may require generating hypotheses and performing complex analysis and connections among texts (WCEPS, 2017).

Webb's Depth of Knowledge categorizes degrees of cognitive complexity one might need to employ when confronted with a question or problem. Webb's DOK contains levels 1 through 4. A Level 1 question or problem is one that presents little cognitive sophistication and requires

only memory recall. Level 1 questions require basic operations that require no analysis of text. (Webb et al., p.70, 2005). Level 2 questions are also deemed to require lower-level thinking in that one need not employ detailed thinking. Level 3 is considered higher because it allows one to do more complex thinking to answer questions such as rationalization and abstract idea recognition. Level 4, the highest Webb Depth of Knowledge level allows students to expand their understanding and learning beyond what is being asked. Much of the Level 4 requirement requires students to expand the knowledge from one text to another and to apply real-life experiences to multiple situations.

This convergent, parallel mixed-methods study was employed to answer the first subquestion: In what way(s) does the language found in the questions on TerraNova English/Language Arts Practice Test for grade 8 compare with the language associated with higher-order thinking found in research literature? One expert coder and I used deductive coding and a double-rater read behind protocol, in accordance with Mayring's (2000) Step Model to categorize each question as per Webb's Depth of Knowledge (DOK). Webb's Depth of Knowledge, with a consideration to Hess's Cognitive Rigor Matrix and Bloom's Taxonomy, was utilized to score each question analyzed in this study as a level 1, 2, 3, or for as prescribed by the DOK.

The double-rater read behind consensus model was used in the study to increase inter-rater reliability and assist the expert coder and researcher with placement of each question after a discussion took place. The expert coder and researcher held six Zoom virtual meetings to calibrate, discuss, and code/score the 220 questions in varied set numbers. Along with the double-rater read behind consensus model, Webb's Alignment Tool and consideration to Hess's Cognitive Rigor Matrix and Bloom's Taxonomy were utilized to ensure reliability.

Conclusions

The results of this convergent, parallel mixed-methods study found that 98% of the test questions are promoting lower level thinking: Level 1 questions accounting for 76%, and Level 2 questions accounting for 22%. The results suggest that the questions promote functional fixedness as opposed to higher-order thinking. Questions categorized as levels 1 and 2 of Webb's DOK are lower level questions with a focus on declarative and procedural knowledge (recall, reconstruction). A consistent focus on lower level declarative and procedural thinking can stunt the complex thinking development of students and lead to functional fixedness. Dunker (1945) described functional fixedness as the phenomenon of one perceiving an entity as having only one function (as cited by Anderson & Johnson, 1966). Functional fixedness is explained by Runco and Chand (1995) as "rigidity or mental set that locks out thinking so that an individual can't see alternatives" (p. 247).

German and Barrett (2005) defined functional fixedness "as a tendency to regard the functions of objects and ideas as fixed" (from Seifert, 2009; p. 190). Seifert (2009) expanded the concept by indicating that functional fixedness may also be referred to as a response set, "the tendency for a person to frame or think about each problem in a series in the same way as the previous problem, even when doing so is not appropriate to later problems" (p. 190). He added that functional fixedness and the response set are obstacles in problem representation, the way that a person understands and organizes information provided in a problem. "If information is misunderstood or used inappropriately, then mistakes are likely - if indeed [a] problem can be solved at all" (p. 190).

In schools, delivery of curriculum and instruction is presented in step-by-step fashion whereby one is to problem solve using specified strategies, method and formulas. This mode of instruction and learning is presented as inspiring critical, creative, or higher-order thinking, yet

functional fixedness is promoted instead. With such delivery of curriculum and instruction, students are virtually denied an opportunity to explore alternate solutions and to think out of the box as they problem solve.

Anderson and Johnson (1966) added that functional fixity includes the concept of Einstellung, which is the challenge of using other techniques to solve an issue after mastering one strategy. The Einstellung effect occurs when prior knowledge obstructs one's ability to reach an optimal and creative solution. In the throes of the Einstellung effect one becomes unable to contemplate additional solutions to a problem, even if one may believe a solution has been found. Arra (2015) noted that the Einstellung effect "leaves us cognitively incapable of differentiating previous experience with the current problem [therefore] we may solve a problem but we don't actually innovate" (p. 1). When consistently taught specified strategies, methods, and formulas to solve problems, "students become unable to think creatively and/or design original solutions to ill-structured problems they are" (Vanderhook, 2020, p. 113).

The concept of analogical transfer is well-established in the literature on higher-order thinking and creativity (Gick & Holyoak, 1980, 1983; Holyoak, 1984, 2005). Psychological studies on creative problem solving have explored factors that determine whether or not one's knowledge about the world or experience with a particular kind of problem or situation can facilitate efforts to solve a new problem with similar features. However, analogical transfer is not always positive. Under certain circumstances, prior knowledge or experience with a particular example or solution strategy may have deleterious effects on one's ability for creative thought (Gentner, 1983; Osman, 2008). Functional fixedness or fixation is an instance of such negative transfer, wherein a solver's experience with a particular function of an object impedes

using the object in a novel way during creative problem solving (Duncker, 1945; Scheerer, 1963).

As noted above, state curriculum policy promotes functional fixedness when state mandated curricula include a majority of content standards with language that is declarative and procedural in nature. State standardized testing policies also promote functional fixedness when assessment of the language of the questions in primarily lower level. Classroom instruction that closely tracks the language of state mandated content standards and standardized tests results in the delivery of curriculum and instruction presented in step-by-step fashion whereby students are compelled to acquire, store, and regurgitate declarative and procedural knowledge. Runco and Chand (1995) investigated how the way in which tasks were presented affected the ways in which students employed higher-order or creative thinking. Students' ability to develop various approaches or think creatively was significantly influenced by the kinds of tasks, activities, and problems presented. The Runco and Chand (1995) study determined that environmental signals, whereby students are able to develop responses through observation or recalling past experiences and prior learning, constrained their capacity to conjure responses exceeding functional fixity. Vanderhook (2020) suggested that the Runco and Chand (1995) study results "suggested that educators should extend their curriculum beyond activities and requirements that have a clearly defined solution or predetermined answer, and instead encourage students to use their problem-solving skills to tackle multiple solutions in real-world situations" (p. 114).

This study found that, of the 220 total questions analyzed and coded, 3 questions were categorized as DOK Level 3, equating to 2% of all questions analyzed and scored in this study. 98% of the questions analyzed required lower level thinking – 76% DOK Level 1 and 22% DOK Level 2 – that did not promote higher-order thinking. In that high stakes testing is a key

determinant of teacher/school effectiveness, it can be said that teaching to the test is a practice widely employed in schools. According to Resnick and Zurawsky (2005), the combination of accountability, the lack of a clear curriculum and cheaper off-the-shelf tests is a recipe for bad teaching. "When teachers match their teaching to what they expect to appear on state [or standardized] tests…students are likely to experience far more facts and routines than conceptual understanding and problem-solving in their curriculum…narrow tests...can become the de facto curriculum" (p. 3).

Jerald (2006) noted that Resnick and Zurawsky do not object to accountability but that they sound a warning that accountability, which can lead to teaching to the test, can lead to inappropriate/lower-level instruction if it is not bolstered with strong curricula and aligned assessments. In *Democracy in Education* (1916) Dewey stated, "No one doubts, theoretically, the importance of fostering in school good habits of thinking" (p. 158). Dewey continues further arguing that "thinking is the method of intelligent learning, of learning that employs and rewards the mind". Thus, "thinking originates in situations where the course of thinking is an actual part of the course of events and is designed to influence the result. The object of thinking is to help reach a conclusion, to project a possible termination on the basis of what is already given...and since the situation in which thinking occurs is a doubtful one, thinking is a process of inquiry, of looking into things, of investigating" (Dewey, 1916, p. 154).

For Dewey then, "all thinking is research, and all research is native, original, with him who carries it on, even if everybody else in the world already is sure of what he is still looking for". In "How We Think", Dewey (1910, p.1) argued that thought can be used loosely to refer to anything that 'comes to mind' or that 'goes through our heads'. In this regard "to think of a thing is just to be conscious of it in any way whatsoever". In this loose sense, thinking signifies

everything that is ‘in our heads’ or that ‘goes through our minds’. Ennis (1964, p. 599) calls this ‘poor critical thinking’. With this, it can be said that the responsibility of educators is to shun inclinations to promote ‘poor critical thinking’ among our students, but instead to strive towards ‘thinking’ in its critical, more reflective sense, or what Ennis calls ‘good critical thinking’.

Dewey (1916) believed that the methods of teaching were cohesive only as measured by the degree to which they focus on producing good thought habits. Standardized testing is a modus operandi in schools and test results are used to measure teacher/school effectiveness. With the importance placed on standardized test results, teaching to the test has become in and of itself a modus operandi; therefore functional fixity is operationalized. Operationalizing lower-level instruction leading to lower-level testing is oppositional to Dewey’s (1916) and Ennis’s (1964) concepts of what promotes good critical thinking.

Webb (1997) categorized skills and concepts necessary for creative/strategic/higher-order thinking and those suited for lower-level thinking. Runco and Chand (1995) noted that procedural knowledge and thinking represents that which is creative/higher-order in nature. Declarative knowledge and thinking represents that which is lower-order – based on facts and recall. Based on the findings of this study, declarative knowledge/thinking was the predominant mode for one to employ in answering the questions on the 8th grade Terra Nova (http://classroom.jc-schools.net/8th/) and the 8th grade IOWA (https://www.ixl.com/standards/iowa/ela/grade-8) Practice Tests in English/Language Arts.

Results of this study revealed that questions on the 8th grade Terra Nova and IOWA Practice Tests in English/Language Arts fall short of both Dewey’s and Webb’s criteria for higher-order/critical thinking. Ninety-eight percent (98%) of the questions contained within the practice tests lacked the cognitive complexity necessary to be categorized beyond a Level 2 of

Webb's DOK. The questions on the practice test questions analyzed in this study also fail to adhere to Dewey's (1916) pronouncement regarding the role of education being to foster habits of good thinking. Instead, the questions found on the practice tests for 8th grade English/Language Arts promote the operationalizing of functional fixedness which, through teaching to tests, serves stymie students from developing and employing higher-order/critical thinking skills and exploring alternative solutions to solving a problem. The multiple-choice format design promotes a myopic mindset in that students need not think beyond the declarative knowledge thought process.

Recommendations for Practice

Findings of this study should serve to alert educators to broaden their breadth and depth of knowledge with regard to higher-order thinking, creative thinking, and cognitive complexity. Such awareness can be found in the literature on the aforementioned topics, study of concepts on thinking espoused by Dewey, and understanding of Webb's Depth of Knowledge, Bloom's Taxonomy and Hess's Cognitive Rigor Matrix. Results of this study indicate that the test companies who published the analyzed practice tests do not offer what they claim vis-à-vis higher-order thinking challenges. Districts should ensure that all stakeholders – administrators, test coordinators, teachers, parents/guardians, and where appropriate, students convene in grade level committees to discuss, review and ultimately adopt tests that challenge students with higher-order thinking problems. Convening the committees by grade level is important as the one-size fits all concept is not a sound practice – different tests may be what is needed by different grade levels.

Both parochial and public schools must bolster professional development that trains teachers and administrators on best practices for promoting higher-order thinking in the delivery

of curriculum and instruction. Professional development should include training such as: Promoting Higher-Order Thinking Skills with Engagement and Motivation (TLC Educational Solutions, 2020). Training should include such topics as Building Metacognition by Developing Understanding of the Learning Process; Opening the Door to Higher-Order Thinking with Strategic Questioning; Scaffolding Instructional Strategies that Raise Student Ability to Operate on Their Own; and Developing Higher-Order Thinking Instructional Approaches. In the Developing Higher-Order Thinking Approaches training, teachers can learn to raise the cognitive demand of learning experiences with strategies that focus on higher-order thinking to help students gain the capacity to process information at a higher level of cognition. This professional development training is steeped in Bloom's Taxonomy and Webb's Depth of Knowledge (DOK) and provides a guide to shifting the delivery of curriculum and instruction toward higher cognitive demand experiences (TLC Educational Solutions, 2020).

The choice of assessment instruments is vital to ensuring students are challenged with higher-order thinking/cognitive complexity questions and problems. Northern and McDougald (2016) researched studies that compared the Depth of Knowledge of ACT Aspire, MCAS, PARCC, and Smarter Balanced tests in relation to the Common Core State Standards found that PARCC tests generally have the highest DOK in ELA/literacy, while ACT Aspire had the highest in mathematics. Their expert review panels found significant variability in the degree to which the four assessments match the distribution of DOK in the Common Core Standards, especially between the grades five and eight assessments for a given program.

A transition into authentic testing is a direction that schools may choose in order to better assess students' higher-order thinking skills. In contrast to traditional assessments, authentic assessments are used to assess what the students can do with the educational material they

have learned, and they aim to evaluate students' abilities in the real-world context (Authentic Assessment Overview, 2001; Bol, Stephenson, O'Connell, & Nunnery, 1998). Authentic forms of assessments encourage students to use their knowledge creatively, and challenge them to express their own interpretations of the material they have learned in class. Unlike traditional assessment, authentic assessment evaluates the accuracy with which a student is able to carry out a function within a given context, and assesses acquired knowledge and higher-order thinking (Sambell & McDowell, 1998; Stiggins, 1997; Viechniki, K. J., Barbour, N., Shaklee, B., Rohrer, J. & Ambrose, R., 1993).

Recommendation for Policy

School leaders must ensure that teachers are properly prepared to implement the aforementioned best practice teaching methods to promote higher-order thinking and cognitive complexity. In efforts to find alternative assessment instruments, school leaders and boards should develop and implement policies that ensure appropriate professional development and convene assessment review by all stakeholders on a consistent basis. Polices should be put in place that sequentially outline steps to be taken vis-à-vis professional development and choice of assessment instrument. Approved policies should include a process for determining the types of professional development opportunities that teachers and administrators are afforded – e.g. staff surveys, research and development teams, strategic planning committees, etc. Processes should be set to occur on a consistent/on-going basis as in three to five-year implementations.

Gentz and Patrick (2016) note the following Policymaker Considerations for Successful Pilot Programs:

- Create pilot programs to catalyze personalized, competency-based learning.
- Fund planning and launch phases.

- Convene practitioners and educators to share best practices through communities of practice.
- Build educator capacity for personalized learning, for developing common performance assessments, and for calibrating and assessing student work and evidence in performance tasks with regard to proficiency and deeper learning.
- Support for the systems change in order to transform to student-centered, competency-based learning.
- Foster a process of continuous improvement, sharing and collaboration (rather than top-down compliance) in monitoring pilots.

School officials and teachers should research, pilot, and implement assessment instruments that are problem-based in nature and that assess higher-order thinking skills such as the CRWA+ test. The Council for Aid to Education (CAE, 2020) noted on their website that **"**faced with unlimited information, students must be able to gather, analyze, and evaluate information effectively. These skills are essential to navigating today's world. The CWRA+ for middle and high schools is a problem-based test that presents real-world, problem-solving performance tasks to assess and measure higher-order/critical-thinking skills. CAE (2020) notes that student data gleaned from CWRA+ administration can help institutions better understand how well students are learning these skills, providing a snapshot of proficiency, growth, and program efficacy" (https://cae.org/flagship-assessments-cla-cwra/why-cla-and-cwra/). The CAE Company claims that they help institutions select and use the best testing model for their particular outcome goals.

Recommendations for Further Research

Further studies should be conducted to examine the types of questions being promoted in the TerraNova and IOWA standardized test instruments as well as other standardized testing

instruments for all grade levels. Such research and study may show different results that could help broaden understanding of the capabilities of testing companies to promote higher-order thinking and cognitive complexity skill development. This study examined questions designed for practice testing in English/Language Arts for grade 8. Further studies should be conducted to analyze and categorize standardized test questions from other grade levels in order that education decision-makers can better understand the overall picture of whether or not the testing companies live up to claims of cognitive complexity.

Examination of other test instruments in various grade levels can serve to ascertain similarities and differences in the frequency and percentage of higher-order questions being asked of students by comparing the questions being asked at different grade levels. Further studies of various norm-referenced standardized tests and investigating authentic learning-based/problem-based tests would help to determine which assessment instrument truly promotes critical thinking through the use of higher-order thinking questions. As noted earlier in this study, a paucity of research exists relative to examining the cognitive complexity/higher-order thinking challenge promoted in the TerraNova and IOWA test instruments. Further studies should be conducted on these tests in various grade levels to determine if there are differences in results as compared to results published in this paper.

This study focused on determining the frequency and percentage of higher-order questions analyzed on the TerraNova and IOWA English/Language Arts Practice Tests for grade 8. The study did not examine the effects of these practice tests, or any other tests on other aspects such as the, student achievement on the New Jersey Student Learning Assessment (NJSLA), Scholastic Aptitude Test (SAT), Preliminary Scholastic Aptitude Test (PSAT), and Advanced Placement exams (AP). Further studies to examine the use of standardized tests and their impact

on student achievement such as, NJSLA, SAT, PSAT, and AP scores, could provide valuable data regarding the overall effectiveness of standardized testing at large. Additional studies can be conducted on the effectiveness of the use of standardized tests in low socioeconomic areas, Abbott districts, or failing districts as it relates to student achievement. Such studies could provide useful information when developing future policies in failing schools/school districts.

There is also a need for future research focused on examining testing companies to determine how great of an influence they have on education policy. Test companies sell assessment products and services that influence schools and the lives of the students they serve. Additionally, further research focused on evaluating curriculum scope and sequence and vertical alignment – specifically targeting how each grade level builds naturally from one to the next as well as a balance of content, teaching/demonstration approaches and literacy standards – would be extremely beneficial for educators. Further review of literature regarding stages of cognitive development and of how much higher-level thinking can be expected at various chronological ages to guide curriculum development and assessment choice at the local level is needed. Qualitative studies of teacher's perceptions about the influence of Webb's DOK and other frameworks such as Hess' Cognitive Rigor Matrix and Bloom's Taxonomy on curriculum development and the promotion of higher-order thinking/cognitive complexity to which students are challenged should provide additional useful information regarding assessment choice.

References

Abosalem, Y. (2016). Assessment techniques and students' higher-order thinking skills. *International Journal of Secondary Education 4*(1), 1-11. Retrieved from http://www.sciencepublishinggroup.com/journal/paperinfo?journalid=193&doi=10.11648/j.ijsedu.20160401.11

Allen, D., Connor, G. L., & De Castro, C. (2017). GEMS: Goals, expectations, measurements and strategies for critical thinking and decision-making. NSU Fischler College of Education. Student Articles 13. Retrieved from https://nsuworks.nova.edu/fse_stuarticles/13

Allen, S. H. L., & Lee, E. W. (2017). Involvement of higher order thinking skills within a preparatory course for the Malaysian university English test. *The English Teacher 40*, 95–112. Retrieved from https://journals.melta.org.my/index.php/tet/article/view/265

Alvermann, D. E., & Phelps, S. F. (1998). Content reading and literacy: Succeeding in today's diverse classrooms, (2nd ed). Boston, MA: Pearson.

American Psychological Association. (2018). APA dictionary of psychology. Washington, DC: American Psychological Association.

Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. OECD Education Working Papers (41). Paris: OECD Publishing. Retrieved from https://static1.squarespace.com/static/5923b7f246c3c49b332b4979/t/59799247d482e9e56c8a822f/1501139531539/21st+Century+Skills+and+Competences+for+New+Millennium+Learners+in+OECD+Countries.pdf

Anderson, L. W., & Krathwohl, D. R., et al. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Boston, MA: Pearson Education Group

Anderson, B., & Johnson, W. (1966). Two kinds of set in problem solving. Psychological Reports. SAGE Journals. Retrieved from https://doi.org/10.2466/pr0.1966.19.3.851

Armstrong, P. (2016). Bloom's taxonomy. Nashville, TN: Vanderbilt University: The Center for Teaching. Retrieved from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Arra, S. (2015). Einstellung effect: What you already know can hurt you. [Web blog post]. Retrieved from https://www.exaptive.com/blog/einstellung-effect

Aungst, G. (2014). Using Webb's depth of knowledge to increase rigor: Categorizing tasks according to the complexity of thought they require is one way for teachers to create a rich learning environment. [Web blog post]. Retrieved from https://www.edutopia.org/blog/webbs-depth-knowledge-increase-rigor-gerald-aungst

Authentic Assessment Overview. (2001). TeacherVision. Sandbox & Co. Retrieved from https://www.teachervision.com/teaching-strategies/authentic-assessment-overview

Ball, A., Joyce, H. D., & Anderson-Butcher, D. (2016). Exploring 21st century skills and learning environments for middle school youth, *International Journal of School Social Work 1*(1). Retrieved from https://doi.org/10.4148/2161-4148.1012

Barahal, S. (2008). Thinking about thinking: Pre-service teachers strengthen their thinking artfully. *Phi Delta Kappan 90*(4), 298-302. Retrieved from https://doi.org/10.1177/003172170809000412

Barak, M., Ben-Chaim, D., & Zoller, U. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education 37*(4), 353-369. Retrieved from DOI: 10.1007/s11165-006-9029-2.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open 2*, 8-14. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352900816000029

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York and Toronto: Longmans, Green.

Bogdan, R. C. & Biklen, S. K. (2016). Qualitative research for education: An introduction to theories and methods (5th ed.). Upper Saddle River, NJ: Pearson.

Bogdanovich, P. (2014). Higher-order questions: Classroom strategy. Dataworks Educational Research. [Web blog post]. Retrieved from https://dataworks-ed.com/blog/2014/10/higher-order-questions/

Boholano, H. B. (2017). Smart social networking: 21st century teaching and learning skills. *Research in Pedagogy 7*(1), 21-29. ERIC Number: EJ1149146. Retrieved from https://eric.ed.gov/?id=EJ1149146

Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. Alexandria, VA: Association for Supervision and Curriculum Development.

Budsankom, P., et al. (2015). Factors affecting higher order thinking skills of students: A meta-analytic structural equation modeling study. Research Paper. *Academic Journals 10*(19), 2639-2652. Retrieved from http://www.academicjournals.org/ERR

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. (*Doctoral dissertation*). Retrieved from https://scholarship.shu.edu/dissertations/2308/

Burrows, B. (2016). What is a standardized test? Article: Study in the USA. [Web blog post]. Retrieved from https://www.studyusa.com/en/a/1284/what-is-a-standardized-test

Carter, N. G. (2015). The quantitative effects of read alouds for primary students Retrieved from https://dspace.carthage.edu/bitstream/handle/123456789/1111/3-25-16The%20Quantitative-Effects-of-Reading-to-Children12%20%281%29%20%282%29.pdf?sequence=1&isAllowed=y

Casner-Lotto, J., & Barrington, L. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce. Research Article. Partnership for 21st Century Skills. ERIC No: ED519465. Retrieved from https://eric.ed.gov/?id=ED519465

Cherry, K. (2019). The 4 stages of cognitive development. Abbott, Inc. Retrieved from https://www.verywellmind.com/piagets-stages-of-cognitive-development-2795457

Collins, R. (2014). Skills for the 21st century: Teaching higher-order thinking. *Curriculum & Leadership Journal 12*(14). Retrieved from http://www.curriculum.edu.au/leader/teaching_higher_order_thinking,37431.html?issueID=12910

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

Council for Aid to Education. (2020). Why CLA+ And CWRA+. Retrieved from https://cae.org/flagship-assessments-cla-cwra/why-cla-and-cwra

Creswell, J. W., & Plano-Clark, V. L. (2011). Designing and conducting mixed methods research. Thousand Oaks, CA: Sage.

Crockett, L. (2015). The critical 21st century skills every student needs to know and why. Retrieved from https://www.academia.edu/11851073/The_Critical_21st_Century_Skills_Every_Student_Needs_and_Why

Dalton, D. W., & Goodrum, D. A. (1991). The Effects of Computer Programming on Problem-Solving Skills and Attitudes. Research Article. SAGE Journal. Retrieved from https://journals.sagepub.com/doi/10.2190/762V-KV6T-D3D1-KDY2

Data Recognition Corporation/DCR. (2017). TerraNova, third edition: Redefining assessment. Retrieved from http://terranova3.com/PDFs/TerraNova_Technical_Quality_Brochure.pdf

David L. (2014). Stage theory of cognitive development (Piaget). Learning Theories. Retrieved from https://www.learning-theories.com/piagets-stage-theory-of-cognitive-development.html

Dede, C. (2009). Comparing frameworks for 21st-century skills. Harvard Graduate School of Education (*Research paper*). Retrieved from http://sttechnology.pbworks.com/f/Dede_(2010)_Comparing%20Frameworks%20for%2021st%20Century%20Skills.pdf

Dewey, J. (1902). The child and the curriculum. The University of Chicago Press, Chicago & London.

Dewey, J. (1910). How we think. Boston, D C Heath & Co.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York: Macmillan.

Dewey, J. (1933). How we think. New York, NY: D. C. Heath and Company. (dissertation). Retrieved from https://scholarship.shu.edu/dissertations/2308/

Dewey, J. (1944). Democracy and education: An introduction to the philosophy of education. New York, NY: The Free Press.

Duncker, K. (1945). On problem-solving (L. S. Lees, Trans.). *Psychological Monographs*, *58*(5), i–113. https://doi.org/10.1037/h0093599. Retrieved from https://psycnet.apa.org/doiLanding?doi=10.1037%2Fh0093599

Eisenman, G., & Payne, B. D. (1997). Effects of the higher-order thinking skills program on at-risk adolescents' self-concept, reading achievement, and thinking skills. Journal Article. *Research in Middle Level Education Quarterly 20*(3). Retrieved from https://www.tandfonline.com/doi/abs/10.1080/10848959.1997.11670099

Elo, S., Kaariainen, M., & Kanste, O. (2014). Qualitative content analysis: A focus on trustworthiness. Research Article. *SAGE Journals 4*(1). Retrieved from https://doi.org/10.1177/2158244014522633

Ennis, R. (1964). A definition of critical thinking. *The Reading Teacher 17*(8), 599-612.

Eriksson, G. I. (1990). Choice and perception of control: The effect of a thinking skills program on the locus of control, self-concept and creativity of gifted students. Gifted Education International. Research Article. *SAGE Journals 6*(3), 135-142. Retrieved from https://doi.org/10.1177/026142949000600303

Fadel, C. (2008). 21st century skills: How can you prepare students for the new global economy? Partnership for 21st Century Skills. PowerPoint. OECD/CERI. Paris, France. Retrieved from https://www.oecd.org/site/educeri21st/40756908.pdf

Ferri-Reed, J. (2013). Leading a multi-generational workforce - Onboarding strategies to supercharge millennial employees. *The Journal for Quality and Participation 36*(2), 32-33. Retrieved from https://www.econbiz.de/Record/leading-a-multi-generational-workforce-onboarding-strategies-to-supercharge-millennial-employees-ferri-reed-jan/10010166725

Fosnot, C. T. (1996). Constructivism: Theory, perspectives, and practice. New York: Teachers College Press.

Francis, E. (2017). What is depth of knowledge? ASCD InService. Retrieved from https://inservice.ascd.org/what-exactly-is-depth-of-knowledge-hint-its-not-a-wheel/

Galvan, J. (2006). Writing literature reviews: a guide for students of the behavioral sciences (3rd ed.). Glendale, CA: Pyrczak Publishing.

Gardner, H. (1993). Creative minds. New York: Basic Books.

Geertsen, H. R. (2003). Rethinking thinking about higher-level thinking. Research Article. *Teaching Sociology 31*(1), 1-19. DOI: 10.2307/3211421. Retrieved from https://www.jstor.org/stable/3211421

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science 7*(2), 155-170. Retrieved from https://doi.org/10.1016/S0364-0213(83)80009-3

Gentz, S., & Patrick, S. (2016). Education innovation pilot programs provide catalyst for localities personalizing learning for k-12 students. AURORA INSTITUTE. [Web blog post] Retrieved from https://aurora-institute.org/blog/education-innovation-pilot-programs-provide-catalyst-for-localities-personalizing-learning-for-k-12-students/

German, T. P., & Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. SAGE Journals. In Seifert, K. (2009), Educational Psychology (2nd ed.). Global Text. The Saylor Foundation. (p. 190). Retrieved from https://www.saylor.org/site/wp content/uploads/2012/06/Educational-Psychology.pdf.

Gibbs, G. R. (2007). Analyzing qualitative data. Chapter 4: Thematic coding and categorizing. Thousand Oaks, CA: Sage. Retrieved from https://dx.doi.org/10.4135/9781849208574.n1

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology 12*(3), 306–355. Retrieved from https://doi.org/10.1016/0010-0285(80)90013-4

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology 15*(1), 1–38. Retrieved from https://doi.org/10.1016/0010-0285(83)90002-6

Grant, C., & Zeichner, K. (1984). Preparing for reflective teaching. Allyn and Bacon: Boston.

Great Schools Partnership. (2014). Standardized Test. The Glossary of Education Reform. Portland, ME. Retrieved from https://www.edglossary.org/standardized-test/

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455. Retrieved from https://doi.org/10.1037/0003-066X.53.4.449

Hand, B., Norton-Meier, L. A., Gunel, M., *et al.* Aligning teaching to learning: A 3-year study examining the embedding of language and argumentation into elementary science classrooms. *International Journal of Science and Math Education 14*, 847–863. Retrieved from https://doi.org/10.1007/s10763-015-9622-9

Heshmat, S. (2013). What is confirmation bias? People are prone to believe what they want to believe. [Web blog post]. Retrieved from https://www.psychologytoday.com/us/blog/science-choice/201504/what-is-confirmation-bias

Hess, K. K. (2004). Applying Webb's depth-of-knowledge (DOK) levels in reading.[online] Retrieved from: http://www. nciea. org/publications. DOKreading_KH08. pdf.

Hess, K. K. (2013). A guide for using Webb's depth of knowledge with common core state standards. Common Core Institute. Retrieved from https://education.ohio.gov/getattachment/Topics/Teaching/Educator-Evaluation-System/How-to-Design-and-Select-Quality-Assessments/Webbs-DOK-Flip-Chart.pdf.aspx

Hess, K. K., Carlock, D., Jones, B., & Walkup, J. R. (2009). What exactly do "fewer, clearer, and higher standards" really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments. Presentation at CCSSO, Detroit, MI, June 2009. Retrieved from http://www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub_date

Hiebert, E., & Mesmer, H. A. E. (2013). Upping the ante of text complexity in the common core state standards: examining its potential impact on young readers. SAGE Journals. (Research article). Retrieved from https://doi.org/10.3102/0013189X12459802

Hmelo, C. E., & Ferrari, M. (1997). The problem-based learning tutorial: Cultivating higher order thinking skills. Journal for the Education of the Gifted. SAGE Journals. Retrieved from https://journals.sagepub.com/doi/10.1177/016235329702000405

Hoffman, N. (2015). Let's get real: Deeper learning and the power of the workplace. Deeper Learning Research Series. (Reference Article). Retrieved from https://luminafoundation.org/files/resources/lets-get-real.pdf

Holyoak, K. J. (1984). Analogical thinking and human intelligence. In: Sternberg R.J., editor. Advances in the psychology of human intelligence. Vol. 2. Hillside, NJ: Erlbaum; pp. 199–230.

Holyoak, K. J. (2005). Analogy. In: Holyoak, K. J. & Morrison, R. G., editors. The Cambridge handbook of thinking and reasoning. Cambridge, UK: Cambridge University Press. Retrieved from https://assets.cambridge.org/97805218/24170/frontmatter/9780521824170_frontmatter.pdf

Jerald, C. D. (2006). Teach to the test? Just say no. The Center for Comprehensive School Reform and Improvement. Washington, DC: Retrieved from https://eric.ed.gov/?id=ED494086

Jirka, S. J., & Hableton, R. K. (2005). Cognitive complexity levels for MCAS assessments. Retrieved from http://www.umass.edu/remp/docs/MCAS-RR-10.pdf

Johnson, P. (2009). The 21st century skills movement. *Educational Leadership 67*(1), 11-11. Retrieved from http://www.ascd.org/publications/educational-leadership/sept09/vol67/num01/The-21st-Century-Skills-Movement.aspx

Jouriles, G. (2014). Here's why we don't need standardized tests. *Education Week 33*(36), 36-40. Retrieved from https://www.edweek.org/ew/articles/2014/07/09/36jouriles.h33.html

Karam, R., Stecher, B. M., Tsai, T., Grimm, G., & Schweig, J. (2016). Mathematics audit of the DoDEA schools. National Defense Research Institute. RAND Corporation. Santa Monica, CA. Retrieved from https://apps.dtic.mil/sti/pdfs/AD1014197.pdf

Kauchak, D. P., & Eggen, P. D. (1998). Learning and teaching: Research-based methods (3rd ed.). Boston: Allyn and Bacon.

Keffe, J. W., et al. (1992). Teaching for thinking. National Association of Secondary School Principals. Reston, VA. Retrieved from https://files.eric.ed.gov/fulltext/ED347647.pdf

Kentucky Department of Education. (2006). Support materials for core content for assessment. Introduction to depth of knowledge (DOK) - Based on Norman Webb's Model (Karin Hess, Center for Assessment/NCIEA, 2005). Retrieved from http://www.e-archives.ky.gov/pubs/Education/cca_mathematics_support_dok.pdf

King, F. J., Rohani, F., & Goodson, L. (1997). Feasibility study: Statewide assessment of listening and verbal communication skills, information literacy skills, and problem-solving skills. Tallahassee: Florida State University. Retrieved from http://www.cala.fsu.edu/files/feasibility_study.pdf

Kohn, A. (2009). Twenty-first century skills? It depends what you mean. [Web blog post]. Retrieved from https://www.alfiekohn.org/blogs/twenty-first-century-skills-depends-mean/

Krulik, S., & Rudnick, J. A. (1993). Reasoning and problem solving: A handbook for elementary school teachers. Allyn and Bacon, Boston.

Kuhn, D. (1999). A developmental model of critical thinking. SAGE Journals. Retrieved from https://doi.org/10.3102/0013189X028002016

Lange, R. T. (2011). Inter-rater reliability. In: Kreutzer J.S., DeLuca J., Caplan B. (eds) Encyclopedia of clinical neuropsychology. Springer: New York, NY

Levy, F., & Murnane, R.J. (2004). The new division of labor: How computers are creating the next job market. Princeton University Press: Princeton, NJ.

Lewis, A., & Smith, D. (1993). Defining higher-order thinking. *Theory Into Practice 32*(3), 131-137. Retrieved from https://doi.org/10.1080/00405849309543588

Bol, L., Stephenson, P. L., O'Connel, A. A., & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research 91*(6), 323-330. Retrieved from https://doi.org/10.1080/00220679809597562

Lombardi, M. M. (2007). Authentic learning for the 21st century: An overview. Educause LearningInitiative. Retrieved from http://www.lmi.ub.edu/cursos/s21/REPOSITORIO/documents/Lombardi_2007

Mayring, P. (2000). Qualitative content analysis. Forum: Qualitative Social Research, p. 40.

McCoog, I. J. (2008). 21st century teaching and learning. Online Submission. Retrieved from https://eric.ed.gov/?id=ED502607

McCutchen, K. L., Jones, M. H., Carbonneau, K. J., & Meuller, C. E. (2016). Mindset and standardized testing over time. *Learning and Individual Differences 45*, 208–213. Retrieved from https://daneshyari.com/article/preview/364521.pdf

McGraw-Hill Communications Team. (2012). CTB/McGraw-Hill helps schools transition to common core assessments through practical guidebook. Retrieved from https://www.mheducation.com/news-media/press-releases/ctbmcgraw-hill-helps-schools-transition-common-core-assessments-through-practical.html

McNeil, L. M. (2000). Contradictions of school reform: Educational costs of standardized testing. Retrieved from https://www.questia.com/library/103329882/contradictions-of-school-reform-educational-costs

Merriam, S. B. (2009). Qualitative research: A guide to design and implementation. San Francisco, CA: Jossey-Bass.

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

Miller, A. (2017). Building classroom dialog using Webb's depth of knowledge. NAEG School of Education. [Web blog post] Retrieved from https://gcci.uconn.edu/2017/11/14/building-classroom-dialogue-using-webbs-depth-of-knowledge/

Mississippi State University. (2009). Webb's depth of knowledge guide: Career and technical education definitions. Retrieved from https://www.aps.edu/sapr/documents/resources/Webbs_DOK_Guide.pdf

Mukeredzi, T. G., Bertram, C., & Christiansen, I. (2018). Investigating teacher learning from a university programme for Foundation Phase teachers. *South African Journal of Childhood Education 8*(1), a524. Retrieved from https:// doi.org/10.4102/sajce. v8i1.524

National Research Council (2011). Assessing 21st century skills: Summary of a workshop. Washington, D.C.: The National Academies Press. Retrieved from https://doi.org?10.17226/13215

Newmann, F. M. (1988). Higher order thinking in the high school curriculum. *NASSP Bulletin 72*(508), 58-64. Retrieved from https://eric.ed.gov/?id=ED304352

Niebling, B .C. (2012). Determining the cognitive complexity of the Iowa core in literacy and mathematics: Implications and applications for curriculum alignment. Midwest Instructional Leadership Council. Iowa Department of Education. Retrieved from

https://iowacore.gov/sites/default/files/determining_the_cognitive_complexity_of_the_iowa_core_in_literacy_and_mathematics_final.pdf

Norris, S. P., & Ennis, R. H. (1989). Evaluating critical thinking: The practitioners' guide to teaching thinking series. Critical Thinking Press and Software: Pacific Grove, CA. Retrieved from https://eric.ed.gov/?id=ED404836

Northern, A. M., & McDougald, V. (2016). How well do next-generation tests measure higher-order thinking skills? Thomas B. Fordham Institute. Retrieved from https://fordhaminstitute.org/national/commentary/how-well-do-next-generation-tests-measure-higher-order-thinking-skills

Oelze, P. (2020). Understanding functional fixedness and how it influences behavior. [Article] Retrieved from https://www.betterhelp.com/advice/psychologists/understanding-functional-fixedness-and-how-it-influences-behavior/

Osman, M. (2008). Positive transfer and negative transfer/antilearning of problem-solving skills. *Journal of Experimental Psychology: General 137*(1), 97–115. Retrieved from https://doi.org/10.1037/0096-3445.137.1.97 P21

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems 25*(6), 493-508. Retrieved from https://doi.org/10.1057/s41303-016-0020-3

Paul, R., & Elder, L. (2006). The miniature guide to critical thinking: Concepts and tools. The Foundation for Critical Thinking. Retrieved from https://www.criticalthinking.org/files/Concepts_Tools.pdf

Paul, R., & Elder, L. (2007). The thinker's guide to analytic thinking: How to take thinking apart and what to look for when you do. The elements of thinking and the standards they must meet. The Foundation for Critical Thinking. Retrieved from https://www.criticalthinking.org/files/SAM_Analytic_Think2007b.pdf

Payne, B. D., & Manning, B. H. (1992). Basal reader instruction: Effects of comprehension monitoring training on reading comprehension, strategy use and attitude. *Reading Research and Instruction 32*(1), 29–38. Retrieved from https://doi.org/10.1080/19388079209558103

Petit, M., & Hess, K. K. (2006). Applying Webb's depth of knowledge and NAEP levels of complexity in mathematics. Karin K. Hess, Center for Assessment. Dover, NH. Retrieved from https://www.rcoe.us/educational-services/files/2013/08/DOK-Math.pdf

Phelps, R. P., et al. (2005). Defending standardized testing. Lawrence Erlbaum Associates (LEA), Publishers. Mahwah, New Jersey, London, xvi. Retrieved from https://www.amazon.com/Defending-Standardized-Testing-Richard-Phelps/dp/0805849122

Piaget, J. (1936). Origins of intelligence in the child. London: Routledge & Kegan Paul.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. Educational Researcher. SAGE Journals. Retrieved from https://doi.org/10.3102/0013189X11405038

Rehmat, A. P. (2015). Engineering the path to higher-order thinking in elementary education: A problem-based learning approach for STEM integration. (*Doctoral Dissertation*). Retrieved form https://digitalscholarship.unlv.edu/thesesdissertations/2497/

Resnick, L., & Zurawsky, C. (2005). Getting back on course: Standards-based reform and accountability. American Educator. Retrieved from https://www.aft.org/periodical/american-educator/spring-2005/getting-back-course

Resnick, L.B. (1987). Education and Learning to Think. National Academy Press. Washington, D.C.

Rimini, M., & Spiezia, V. (2016). Skills for a digital world. Directorate for Economic Co-operation and Development/OECD. Draft Background Paper for Ministerial Panel 4.2. Retrieved from https://one.oecd.org/document/DSTI/ICCP/IIS(2015)10/REV2/en/pdf

Runco, M. A., & Chand, I. (1995) Cognition and Creativity. *Educational Psychology Review 7*, 243-267. Retrieved from http://dx.doi.org/10.1007/BF02213373

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education 23*(4), 391-402. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/0260293980230406

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 https://files.eric.ed.gov/fulltext/ED536959.pdf

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology Archive 35*, 31-38. Retrieved from https://www.semanticscholar.org/paper/Problem-Based-Learning%3A-An-Instructional-Model-and-Savery-Duffy/d5d150c7126c0585cebc6888b2a00aaa4f232bac

Sawchuk, S. (2009). Backers of 21st-century skills take flak. *Education Week 28*(23), 1. Retrieved from https://eric.ed.gov/?id=EJ831327

Schafersman, S. (1991) An introduction to critical thinking. Retrieved from https://facultycenter.ischool.syr.edu/wp-content/uploads/2012/02/Critical-Thinking.pdf

Schafersman, S. (1997) An introduction to science: Scientific thinking and the scientific method. Retrieved from: https://www.geo.sunysb.edu/esp/files/scientific-method.html

Scheerer, M. (1963). Problem-solving. *Scientific American, 208*(4), 118-128. Retrieved from https://psycnet.apa.org/record/1964-03721-001

Schulz, H., & FitzPatrick, B. (2016). Teachers' understandings of critical and higher order thinking and what this means for their teaching and assessments. *Alberta Journal of Educational Research 62*(1), 61-86. Retrieved from https://journalhosting.ucalgary.ca/index.php/ajer/article/view/56168/pdf

Seifert, K. (2009). Educational Psychology (2nd ed.). Global Text. The Saylor Foundation. https://www.saylor.org/site/wp-content/uploads/2012/06/Educational-Psychology.pdf.

Sforza, D., Tienken, C. H., & Kim, E. (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), 4-30. Retrieved from https://eric.ed.gov/?id=EJ1090309

Silva, E. (2009). Measuring Skills for 21st-Century Learning. *Phi Delta Kappan 90*(9), 630–634. Retrieved from https://doi.org/10.1177/003172170909000905

Sizemore, J. M. (2015). Intentional depth of knowledge and its effects on k-12 student engagement. Bellarmine University Graduate Theses, Dissertations, and Capstones. Retrieved from https://scholarworks.bellarmine.edu/cgi/viewcontent.cgi?article=1019&context=tdc

Smith, F. (1992). To think: In language, learning, and education. Routledge: London.

Snowman, J., & McCowan, R. (2015). Psychology applied to teaching. Cengage Learning. Stamford, CT.

State of New Jersey Department of Education. (2017). New Jersey Student Learning Standards: Standard 9 – 21st century life and careers. Retrieved from http://www.state.nj.us/education/aps/cccs/career/

Sternberg, R. J. (2005). WICS: A model of positive educational leadership comprising wisdom, intelligence, and creativity synthesized. *Educational Psychology Review 17*(3), 191-262. Retrieved from https://www.psycholosphere.com/WICS%20-%20A%20model%20of%20positive%20educational%20leadershipcomprising%20wisdom,%20intelligence,%20and%20creativity%20by%20Sternberg%20(1).pdf

Stiggins, R.J. (1994). Student-centered classroom assessment. Retrieved from https://www.semanticscholar.org/paper/Student-centered-classroom-assessment-Stiggins/d44ed1ed460c5ef89a4dda59593f49ecdeadbc0a

Sydoruk, P. D. (2018). An analysis of the higher order thinking requirements of a grade 8 online-based english language arts skills program. Seton Hall University Dissertations and Theses (ETDs). Retrieved from https://scholarship.shu.edu/dissertations/2495

Tan, S. Y., & Halili, S. H. (2015). Effective teaching of higher-order thinking (HOT) in education. *The Online Journal of Distance Education and E-Learning 3*(2), 41–47. Retrieved from https://tojdel.net/journals/tojdel/articles/v03i02/v03i02-04.pdf

Tankersley, K. (2005). Literacy strategies for grades 4-12: Reinforcing the threads of reading. ASCD. Alexandria: VA.

Tanner, D. (2016). Jean Piaget's debt to John Dewey. *AASA Journal of Scholarship and Practice 13*(1), 9–25. Retrieved from https://www.aasa.org/uploadedFiles/Publications/JPS-Spring2016.FINAL.v2.pdf

Tanner, D., & Tanner, L. (2007). Curriculum development: Theory into practice. Pearson: Upper Saddle River, NJ.

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Commun. Assoc. Inf. Syst. 37*(6). Retrieved from https://aisel.aisnet.org/cais/vol37/iss1/6/

Thompson, L. (2013). Measuring up with the TerraNova 3 ACSI achievement test . *CSE Magazine 16*(1). Retrieved from https://www.acsi.org/resources/cse/cse-magazine/measuring-up-with-the-terranova-acsi-achievement-test-161

Tienken, C. H. (2013). Defying standardization: Creating curriculum for an uncertain future. Rowman & Littlefield Publishing Group. Lanham, Maryland.

Tienken, C. H., & Orlich, D. C. (2013). The school reform landscape: Fraud, myth, and lies. Rowman and Littlefield, Lanham, MD.

Tienken, C. H., & Zhao, Y. (2010). Common core national curriculum standards: More questions…and answers. *AASA Journal of Scholarship and Practice 6*(4), 3-13. Retrieved from https://www.aasa.org/uploadedFiles/Publications/Journals/AASA_Journal_of_Scholarship_and_Practice/Winter_10_FINAL.pdf

TLC Educational Solutions. (2020). Professional development for educators. Retrieved from http://tlceducationalsolutions.com/

U.S. Department of Education Institute of Education Sciences National Center for Education Statistics. (2013). Testing integrity symposium: Issues and recommendations for best practice. Retrieved from https://nces.ed.gov/pubs2013/2013454.pdf

Vanderhook, C. A. (2020). The type of questions being promoted in a 10th grade social studies textbook. Seton Hall University Dissertations and Theses (ETDs). Retrieved from https://scholarship.shu.edu/dissertations/2730

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior 72*, 577-588. Retrieved from https://doi.org/10.1016/j.chb.2017.03.010

Viechniki, K. J., Barbour, N., Shaklee, B., Rohrer, J., & Ambrose, R. (1993). The impact of portfolio assessment on teacher classroom activities. *Journal of Teacher Education 44*(5), 371-377. Retrieved from https://doi.org/10.1177/0022487193044005007

Watson, S. (2020). Higher-order thinking skills (HOTS) in education. Retrieved from https://www.thoughtco.com/higher-order-thinking-skills-hots-education-3111297

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (NISE Research Monograph No. 6). Retrieved from https://eric.ed.gov/?id=ED414305

Webb, N. L. (2002). Depth-of-knowledge levels for four content areas. Retrieved from https://tapconyc.weebly.com/uploads/1/9/1/5/191529/depth_of_knowledgelevels.pdf

Webb, N., Alt, M., Ely, R., & Vesperman, B. (2005). Web Alignment Tool (WAT) training manual. Wisconsin Center for Educational Research *1*(1) 1–177.

Welch, C., & Dunbar, S. (2014). Measuring growth with the IOWA assessments. Retrieved from https://itp.education.uiowa.edu/ia/Research.aspx

Welch, C., Dunbar, S., & Rickels, H. (2016). Measuring student growth with the IOWA assessments. ITP Research Series. Retrieved from https://itp.education.uiowa.edu/ia/documents/Measuring-Student-Growth-in-Iowa-with-the-Iowa-Assessments.pdf

Wen Q. F., Wang J. Q., Zhao C. R., Liu Y. P., & Wang H. M. (2009). Constructing the theoretical framework of Chinese foreign language college students' critical thinking ability measuring tool. *Foreign Language Circles 1*, 37-43. Retrieved from http://www.academypublication.com/ojs/index.php/tpls/article/view/tpls0803349353

Wisconsin Center for Education Products And Services, Inc.: WCEPS. (2017). WebbAlign. Retrieved from https://dpi.wi.gov/sites/default/files/imce/social-studies/pdf/Social_Studies_-_WCEPS_WebbAlign_Webbs_DOK_Summary.pdf

Yen, T. S., & Halili, S. H. (2015). Effective teaching of higher-order thinking (HOT) in education. *The Online Journal of Distance Education and e-Learning 3*(3), 41-45. Retrieved from https://pdfslide.net/documents/effective-teaching-of-higher-order-thinking-shin-yen-siti-hajar-halili-ajai912gmailcom.html

Yuan, K., & Le, V. (2012). Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests. [Working paper]. Retrieved from https://www.rand.org/content/dam/rand/pubs/working_papers/2012/RAND_WR967.pdf

Zaidieh, A. J. (2012). The use of social networking in education: Challenges and opportunities. *World of Computer Science and Information Technology Journal (WCSIT) 2*(1), 18-21. Retrieved from https://pdfs.semanticscholar.org/76e2/1d0c5cc14238463a09eec33d5d06573a32d2.pdf

Zhong Q. Q. (2002). Prospect of critical thinking and teaching. Global Education, 1, 34-38. In Zhou, Z. (2008). A study on the cultivation of critical thinking ability of English majors. *Theory and Practice in Language Studies 8*(3), 349-353. Retrieved from https://www.academypublication.com/ojs/index.php/tpls/article/view/tpls0803349353

Zhou, Z. (2018). A study on the cultivation of critical thinking ability of English majors. *Theory and Practice in Language Studies 8*(3), 349-353. Retrieved from https://www.academypublication.com/ojs/index.php/tpls/article/view/tpls0803349353

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low achieving students: Are they mutually exclusive? *Journal of the Learning Sciences 12*(3) 145-181. Retrieved from https://www.jstor.org/stable/1466891?seq=1

APPENDIX

Section C. Reading DOK Levels

In language arts, four DOK levels were used to judge both reading and writing objectives and assessment tasks. The reading levels are based on Valencia and Wixson (2000, pp. 909–935). The writing levels were developed by Marshá Horton, Sharon O'Neal, and Phoebe Winter.

Reading Level 1. Level 1 requires students to receive or recite facts or to use simple skills or abilities. Oral reading that does not include analysis of the text, as well as basic comprehension of a text, is included. Items require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Support ideas by reference to verbatim or only slightly paraphrased details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage.

Reading Level 2. Level 2 includes 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. Some important concepts are covered, but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a reading selection.
- Identify and summarize the major events in a narrative.

Reading Level 3. Deep knowledge becomes a greater focus at Level 3. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Standards and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Explain or recognize how the author's purpose affects the interpretation of a reading selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

Reading Level 4. Higher-order thinking is central and knowledge is deep at Level 4. The standard or assessment item at this level will probably be an extended activity, with extended time provided for completing it. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require the application of significant conceptual understanding and higher-order thinking. Students take information from at least one passage of a text and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts. Some examples that represent, but do not constitute all of, Level 4 performance are:

- Analyze and synthesize information from multiple sources.
- Examine and explain alternative perspectives across a variety of sources.
- Describe and illustrate how common themes are found across texts from different cultures.

Writing Level 1. Level 1 requires the student to write or recite simple facts. The focus of this writing or recitation is not on complex synthesis or analysis, but on basic ideas. The students are asked to list ideas or words, as in a brainstorming activity, prior to written composition; are engaged in a simple spelling or vocabulary assessment; or are asked to write simple sentences. Students are expected to write, speak, and edit using the conventions of Standard English. This includes using appropriate grammar, punctuation, capitalization, and spelling. Students demonstrate a basic understanding and appropriate use of such reference materials as a dictionary, thesaurus, or Web site. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Use punctuation marks correctly.
- Identify Standard English grammatical structures, including the correct use of verb tenses.

Writing Level 2. Level 2 requires some mental processing. At this level, students are engaged in first-draft writing or brief extemporaneous speaking for a limited number of purposes and audiences. Students are expected to begin connecting ideas, using a simple organizational structure. For example, students may be engaged in note-taking, outlining, or simple summaries. Text may be limited to one paragraph. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Construct or edit compound or complex sentences, with attention to correct use of phrases and clauses.
- Use simple organizational strategies to structure written work.
- Write summaries that contain the main idea of the reading selection and pertinent details.

Writing Level 3. Level 3 requires some higher-level mental processing. Students are engaged in developing compositions that include multiple paragraphs. These compositions may include complex sentence structure and may demonstrate some synthesis and analysis. Students show awareness of their audience and purpose through focus, organization, and the use of appropriate compositional elements. The use of appropriate compositional elements includes such things as addressing chronological order in a narrative, or including supporting facts and details in an informational report. At this stage, students are engaged in editing and revising to improve the quality of the composition. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Support ideas with details and examples.
- Use voice appropriate to the purpose and audience.
- Edit writing to produce a logical progression of ideas.

Writing Level 4. Higher-level thinking is central to Level 4. The standard at this level is a multi-paragraph composition that demonstrates the ability to synthesize and analyze complex ideas or themes. There is evidence of a deep awareness of purpose and audience. For example, informational papers include hypotheses and supporting evidence. Students are expected to create compositions that demonstrate a distinct voice and that stimulate the reader or listener to consider new perspectives on the addressed ideas and themes. An example that represents, but does not constitute all of, Level 4 performance is:

- Write an analysis of two selections, identifying the common theme and generating a purpose that is appropriate for both.

Examples Applied to Objectives and Assessment Items

Sample Language Arts Objectives

Use the language arts DOK levels on the previous pages to determine the DOK levels for the following five sample objectives. When you are finished, turn the page to see whether you agree with the way we coded these objectives! After this, try using the DOK levels on the Sample Language Arts items.

Objective 1. Identify cause and effect, and understand main idea and purpose implied by text.

Objective 2. Recall elements and details of story structure, such as sequence of events, character, plot, and setting.

Webb, N., Alt, M., Ely, R., & Vesperman, B. (2005). Web Alignment Tool (WAT) training manual. Wisconsin Center for Educational Research *1*(1) 1–177.