Fall 10-21-2021

An Analysis of Higher-Order Thinking Requirement of the 2018 PSAT/NMSQT

Ryan Esdale
ryan.esdale@student.shu.edu

An Analysis of Higher-Order Thinking Requirement of the 2018 PSAT/NMSQT

Ryan W. Esdale

Dissertation Committee

Eloise Marks-Stewart, EdD
David Reid, PhD
Michael Thumm, EdD

Submitted in partial fulfillment of the requirements for the degree of
Doctor in Education
Department of Educational Leadership, Management, and Policy

Seton Hall University
2021

SETON HALL UNIVERSITY

1856

COLLEGE OF EDUCATION & HUMAN SERVICES

DEPARTMENT OF EDUCATION LEADERSHIP MANAGEMENT & POLICY

APPROVAL FOR SUCCESSFUL DEFENSE

Ryan Esdale has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Fall** Semester.

DISSERTATION COMMITTEE
(please sign and date)

Dr. Eloise Marks-Stewart	10.21.21
Mentor	**Date**
Dr. David Reid	10.21.21
Committee Member	**Date**
Dr. Michael Thumm	10.21.21
Committee Member	**Date**

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

ABSTRACT

As educators and school leaders work towards building students' capacity throughout a student's academic career, the goal is to facilitate opportunities for students to develop college and career readiness skills that can be applied to novel and challenging problems long after high school graduation. Hess' Cognitive Rigor Matrix (HCRM) is a tool that objectively and effectively measures prompts, learning exercises, or test questions by blending the Revised Bloom's Taxonomy and Webb's Depth of Knowledge. Thus, a learning environment that cultivates higher-order thinking skills can be accurately examined and provide educators, researchers, and school leaders with data to make more informed decisions about delivering instruction and proficiently assessing the students' college and career readiness. Increasingly, the culture of education in the United States has been impacted by the results of high-stakes standardized testing. Essentially, the significance of the Preliminary Scholastic Aptitude Test and National Merit Scholarship Qualifying Test (PSAT/NMSQT) has grown to influence a student's academic trajectory, a high school's evaluation. The PSAT/NMSQT has also contributed to a standardized testing industry and test-prep industry that has blossomed into the billion-dollar range. Moreover, the PSAT/NMSQT purportedly measures a student's college and career readiness. Given the vast scope and depth of consequences that accompany students' PSAT/NMSQT scores, the exam's claims to examine career and college readiness skills were thoroughly investigated through the lens of qualitative document analysis. The study was able to identify and quantify the higher-order thinking skills of the 2018 PSAT/NMSQT assessed on the examination using subject-specific HCRMs.

This research has added an objective measure of the language used on the 2018 PSAT/NMSQT and found that the extent to which higher-order thinking is being used on the exam does not support its stated purpose of assessing college readiness. The researchers methodically and systematically reviewed the language of each question on the 2018 PSAT/NMSQT through the lens of subject-specific HCRMs to determine the overall level of cognitive rigor of the 2018 PSAT/NMSQT and the level of cognitive rigor of each section (Reading, Writing, and Math). Overall, 31.6% of the questions in the 2018 PSAT/NMSQT contained language compared to the specified HCRMs definition of higher-order thinking. Significantly, most questions (51.8%) on the 2018 PSAT/NMSQT required test-takers to simply apply skills and concepts to solve the test question. The study concluded that the PSAT/NMSQT should look to enhance the math section in order to improve the cognitive rigor required to complete the test question. Additionally, local education agencies should create more effective learning experiences for their students to practice higher-order thinking skills rather than relying on a high-stakes standardized test. Lastly, Hess' Cognitive Rigor Matrix is a well-acquitted, fully researched measurement tool that can be applied to local assessments and develop a culture of deeper learning through a validated, yet differentiated catalog of data-driven assessments that inspires students to engage in challenging learning experiences.

Keywords: Hess' Cognitive Rigor Matrix, HCRM, Cognitive Rigor, higher-order thinking skills, college and career readiness, high stakes standardized testing, PSAT, PSAT/NMSQT

ACKNOWLEDGEMENTS

The construction of this dissertation and path toward a Doctoral Degree from Seton Hall University has seen long nights, early mornings, a little bit of grit, and an incredible amount of support from family and friends. In May of 2017, I found myself in Ramsey School District's Board of Education office in a second-round interview with Dr. Murphy, Dr. Dinning, and Dr. Thumm. I was the only person in the room who had not earned a Doctoral Degree. Not only was I in awe of their command of all things in the field of education, but even more impressive was their genuine amicability, inspiring dignity, and effortless grace. I could not shake a feeling of self-loathing inadequacy sinking my heart to the floor (not by anything they said or did, quite the opposite, the wave of emotion was more of a feeling that I did not deserve to have a seat at the table). I hated that feeling. I said to myself, that will never happen to me again. As time went on, I was so incredibly lucky to be granted a school leadership position at Ramsey High School, and Dr. Murphy, Dr. Dinning, and Dr. Thumm continue to inspire me each and every single day as I learned that it is not three little letters attached to your last name, but how hard you work for students and the learning community that is most significant. I was told early on in my career at Ramsey High School that we stand on the shoulders of giants. Dr. Murphy, Dr. Dinning, and Dr. Thumm… thank you for allowing me to stand on your shoulders as you have shown me the way.

To Holly Falcone, always implicitly and explicitly demanding greatness... I thank you for your motivation and pushing me forward. To Melissa Cruz, for her selfless dedication to opening the field of mathematics to all learners and being a contributor to the evaluation of the math sections. To Carmen Lacherza, for your friendship, your mission to help all students

become readers, and your incredible help in reviewing the Reading and the Writing sections.

To the framers of this dissertation, Dr. Sydoruk, Dr. Fitzhugh, Dr. Tienken, Dr. Katz, and Dr. Corino. You may never know the impact of the time, interest, and development you have given me during my pursuit of higher education. Thank you for giving me the blueprint! Dr. Marks-Stewart, my second mentor. Your kindness, patience, and constructive feedback were extremely helpful. Picking up where Dr. Gibbs-Thomas left off was no small challenge. You handled the transition with tender care, unconditional positive regard, and unparalleled gregariousness. Thank you for the confidence you instilled in me. To Dr. Reid, without your friendship, collegiality, and coaching, this completed dissertation would not be even remotely possible. Seton Hall University is a better place because of your leadership and your commitment to excellence.

To Dr. Thumm, words cannot express how much you have helped me. Your influence on my dissertation, my career, and my nascent school leadership skills has been outstanding. Getting your Doctorate degree at such a young age (33) was a goal that I constantly chased, and helped push me through some challenging obstacles.

This dissertation is devoted to my family. My mother and father, words cannot express how thankful I am for your ability to always lift me up. No matter the day, no matter the situation, you are always there to lend a kind word, share a unique perspective, and demand perseverance through motivating conversations. The summer of 2020 will fondly be remembered for toiling away on the Chromebook on the back deck for hours at a time, being present in the colloquial chatter abound, but also focused on creating my little universe of higher-order thinking skills and cognitive rigor. My sister, Arielle, your genuine interest in my progress was thoroughly appreciated! Even when the conversation about the approach to analyzing the PSAT/NMSQT got a little wonky, you still grasped the idea and grappled with understanding the nuance of the conversation. Oh, one more thing… I win every debate now.

Finally, to my beautiful wife Jenna, my daughter, Siena, and my son, Roman. Jenna, the burden you carried throughout my pursuit of higher education is incalculable. The weekends alone, the caring for children alone, the nap times that were spent tapping away on a Chromebook instead of sitting next to you… I am forever indebted to you for your patience, endurance, independence, confidence in your ability to parent singularly, and your unwavering support. This would not and could not have been done without you. Siena, I worked and worked and worked to ensure that you would never see me typing or reading a book. I never wanted you to feel like you were second, and I never wanted you to know of a dissatisfying moment where your father was too busy to run, play, laugh, swim, and be present with you. You may be too young to understand now, but I sincerely hope that in time you can fathom the joy you have brought me and how the choice to work on this dissertation at 11:00 pm or 5:00 am was inexorably worth every minute I got to spend with you. I cannot wait to see what is in store for you - your energy, laughter, kindness, curiosity, and your infectious enthusiasm inspire me every day. The world is not ready for you, Champ. Roman, this dissertation's journey arrived at its conclusion as you were just about six months old. Your experience with me working on this dissertation is more of an interminable message than an interaction. Different people have tried to place me in different boxes with different labels. I do not want that for you, at all. Please use this dissertation as an example. No matter what the future holds for you, you must have the fortitude to boldly choose your path, define your own life without fear of condemnation or judgment. Go confidently wherever you want to go, and attack every second of every day with a work ethic that is second to none.

Lastly, this dissertation is dedicated to all those who have been told that they can't or shouldn't...

"If you can't fly then run, if you can't run then walk, if you can't walk then crawl, but whatever you do you have to keep moving forward."
-- Martin Luther King, Jr.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

Background

In the book *Deeper Competency Based Learning: Making Equitable, Student-Centered, Sustainable Shifts,* authors Hess, Colby, and Joseph (2020) state:

> As a nation, we have simply outgrown the one-size-fits all instructional and assessment practices of the past. Today, schools are embracing new models of teaching and learning that personalize a child's education, require evidence of proficiency, and lead to graduation from high school with future readiness for the workplace and further education.

To that end, *Teaching for Deeper Learning: Tools to Engage Students in Meaning Making*, authors McTighe and Silver (2020) wrote:

> The ultimate goal is transfer - that is, empowering students to become proficient enough users of the tools that they'll be able to apply these tools (and the embedded thinking skills) independently in future learning situations.

With this in mind, the purpose of this dissertation is to identify the higher-order thinking skills the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT) assesses on their examination. This study aims to shed light on the standardized testing system that runs concurrently with local, state, and national education policy and practice regarding career and college readiness. The term 21st-century skills have been applied to multiple facets of education. No Child Left Behind resulted in students, families, parents, guardians, teachers, and administrators discussing and grappling with the idea of teaching to a standardized test. Many fear that if the U.S. education system does not produce better-educated students, especially in the domains of science and math, then there is an increased risk of jobs being taken by better-educated students in other countries (Wagner, 2008).

Increased standardized testing has affected the delivery of instruction in classrooms and the primary lens by which students are measured against one another. To that end, Wagner and Dintersmith (2015) posit that “We have been conditioned to view some types of credentials as being marks of outsized distinction… Education credentials are our country’s caste system”. In essence, the modern-day caste system that post-secondary institutions have engineered sadly reduces dynamic and diverse teenagers to numbers on a piece of paper: grade point average and test scores. The college entrance exam has become a make-or-break rite of passage in an American and global society that now values credentials over-application of knowledge through the use of college and career readiness skills. Therefore, the outcomes of the standardized tests determine future paths, such as rigorous coursework in high school, access to Advanced Placement (AP) courses, graduating high school, and admission to rigorous post-graduate institutions. Accordingly, a comprehensive analysis of what exactly is being tested on each question of the PSAT/NMSQT, which is, in essence, a gatekeeper to certain education and career paths, is required to serve our students and community of educators.

The standardized testing of high school students in New Jersey has become more high stakes than ever for two main outcomes: high school graduation proficiency and college admissions. Essentially, the New Jersey Department of Education (NJDOE) annually publishes a large amount of data regarding school performance to present a comprehensive picture of overall school performance, and the College and Career Readiness section provides information about how student behaviors correlate with greater success in college and careers (NJDOE, 2019). Subsequently, the Scholastic Aptitude Test (SAT), American College Testing (ACT), and the Preliminary Scholastic Aptitude Test and National Merit Scholarship Qualifying Test

(PSAT/NMSQT) are the primary tools used to gather data regarding college and career readiness for the NJDOE. Additionally, New Jersey Monthly Magazine is a publication that constructs a metric every two years in which schools are arranged based on test scores and college placement. Naturally, with multiple organizations utilizing the data of the PSAT/NMSQT to measure high school students' abilities, it is more important than ever to review the validity of the PSAT/NMSQT as a measurement of higher-order thinking skills.

Problem Statement

The purpose of this dissertation is to identify the higher-order thinking skills that the PSAT/NMSQT actually assesses. Each PSAT/NMSQT question requires a certain set of skills and knowledge to complete the prompt or answer the question correctly. What exactly are the skills required? What is the depth of knowledge needed? Does the question require transfer of knowledge or rote memorization? Can the question assess critical thinking skills and creativity? Can a student be a stellar communicator but have PSAT/NMSQT scores that illustrate those abilities? This study focuses on analyzing each question of the PSAT/NMSQT in a measured, objective, and thorough fashion to identify what skills each question requires from the test-taker.

A critical outgrowth of progressive education in modernity illustrates the dichotomy between instructional delivery and assessment. For example, the "Portrait of a Graduate" (Battelle, 2019) initiative encourages high schools to create authentic learning experiences for students based on a localized set of competencies serving as the shepherds of all learning exercises. However, the evaluation of high schools in the state of New Jersey hinges upon basic statistics like graduation rate, 5-year graduation rate, and whether or not students are college and career-ready. Quite simply, the data on the college and career readiness portion of the NJDOE School Performance Report (*Table 1*) contains only the following:

Table 1. College and career readiness portion of the NJDOE School Performance Report (2018).

PSAT/NMSQT, SAT & ACT - Participation
PSAT/NMSQT, SAT, & ACT Performance
Advanced Placement (AP) / International Baccalaureate (IB) Courses Offered
AP/IB Coursework - Participation and Performance
Dual Enrollment Coursework - Participation
Career and Technical Education Participation
Structured Learning Experiences Participation
Industry Valued Credentials
Industry Valued Credentials by Career Cluster
Mathematics Course - Participation
Science Course - Participation
Social Studies and History - Course Participation
World Languages - Course Participation
Seal of Biliteracy
Visual and Performing Arts - Course Participation

In particular, the New Jersey Performance Report Reference Guide details the nature of the NJDOE's evaluation as part of New Jersey's accountability system. The College and Career Readiness section of the New Jersey School Performance Reports provides information about student behaviors that correlate with greater success in college and career. These behaviors include taking college entrance exams, taking advanced coursework, participating in visual and performing arts courses, and participating in career and technical education programs.

In a parallel evaluative process, the New Jersey School Performance report also utilizes data housed in the PSAT/NMSQT exam results. Data collected from each school district on the PSAT/NMSQT is limited to the PSAT/NMSQT 10 and PSAT/NMSQT exams and does not include performance on the PSAT/NMSQT 8/9 exam. Inevitably, the growth of students taking a formative exam is not taken into consideration as students matriculate from 8th grade to 11th grade. Therefore, PSAT/NMSQT participation and performance is based on tests taken during the reported school year. The data collected on the PSAT/NMSQT is combined with the data from the SAT and ACT, which are both used as college admissions examinations, to illustrate the school's preparation of students for college and career readiness. Subsequently, the SAT, ACT, and the PSAT/NMSQT are the main tools used to judge college and career readiness in the eyes of college admissions counselors and the NJDOE. Moreover, New Jersey Monthly Magazine is a publication that constructs a ranking every two years in which schools are arranged based on test scores and college placement solely on the NJDOE Annual School Performance Report Card. The ranking includes specific metrics such as percentage of students scoring at or above 530 for the math SAT and percentage of students scoring at or above 480 for reading and writing SAT. To that end, the need to establish norms and trends of students who contain the higher-order thinking skills for college and career readiness should be judged based on assessments that have been validated as accurate tools; currently used standardized tests such as the PSAT/NMSQT are not well validated. For example, Harlow (2019) questions the real-world utility of standardized tests by stating that:

> Many in the field of cognitive science believe these exams shouldn't matter in the first place – moment-in-time evaluations are fraught with problems and don't provide an accurate view of real knowledge or potential.

Simply put, given the vast scope and depth of consequences that accompany students' PSAT/NMSQT scores, the exam's claims to examine career and college readiness skills must be more thoroughly investigated.

In a related concern regarding standardized testing, there are also socio-economic issues focusing on the millions of dollars being spent on standardized testing. Invariably, school districts spend an estimated $68,000,000 on the PSAT/NMSQT each year. Shockingly, the College Board itself is a non-profit company that had $1,067,701,847 in total revenue in 2017 (Tigas, 2013). In addition, the fairness of admissions testing has increasingly been called into question, the harshest criticisms often focusing on the outsized advantage the test prep industry gives students whose families are able to pay hundreds to thousands of dollars for test prep with the hopes of sky-high scores. The test prep industry, which includes informal tutoring, group and online classes, books, and mobile apps, has been valued anywhere between $1 billion and $4 billion" (Howard, 2014). Astoundingly,

> This industry actively lobbies federal and state legislatures. With billions of dollars on the line, there are powerful, entrenched interests arguing that, in fact, miniscule differences in test scores should be reported, craved, and chased, costing millions of families time, money, and opportunity. (Dintersmith & Wagner, 2015)

To illustrate, the College Board admitted that SAT coaching could boost scores significantly as some services ask for $1,000 to $2,500 from students and their families (Strauss, 2017).

Additionally, a host of well-funded startups entered the space, promising adaptive, customizable learning through technology over the past decade. The most notable were Knewton ($157m raised), Grockit ($45m raised), and Magoosh. But Magoosh, which raised less than a million dollars in funding, emerged as the largest online test prep company and has been cash-flow positive since 2012 (Vincent, 2018). Inevitably, the quick pivot to disrupt a test prep industry reveals the incredible amount of profiteering off of the standardized testing of high school students. Consequently, the incredibly large financial windfalls surrounding College Board's standardized testing further compels a study to investigate the validity of the claims made by organizations such as College Board that they can pinpoint the strengths and weaknesses of students preparing for college and careers in the 21st century.

Purpose Statement

This qualitative document analysis focused on comparing, analyzing, and describing the language of complex thinking embedded within the PSAT/NMSQT. The cascading issue that lies ahead of the potential overvaluing of test scores is that there is a disconnect between how to teach and check for understanding of the utility of skills and application of knowledge. For example, in Brookhart's book, *How to Assess Higher-Order Thinking Skills in Your Classroom* (2020); "definitions that I find helpful fall 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." Can a multiple choice PSAT/NMSQT administration over the course of three hours measure those three categories of higher-order thinking skills? Specifically, Hess, Colby, and Joseph (2020) discuss how educators should manufacture opportunities to sharpen higher-order thinking skills (analysis, prototyping, synthesis, experimentation, creating an argument) that promote student success beyond the classroom in which students gain a deep understanding of broad topics to eventually equip the students with applicable knowledge and skills in preparation for career, college, and life challenges. The expectations set by an educator of a student should be to master the skills taught in the classroom and then couple those skills with a growth mindset as a springboard to engage, grapple with, and overcome new and open-ended, complex problems (Hess, Colby, & Joseph, 2020).

A recent New York Times article underscores the burden of bored, disengaged, bewildered, and disaffected students at high schools across the U.S., in which the article echoed the sentiment stated by Brookhart (2020) in terms of the dire need for students being able to engage in open-ended, real-world problems:

> More radically, what was powerful about extracurriculars is that students were supported in leading their learning. They were taking responsibility for teaching others and gradually becoming the ones who upheld the standards of the field. The more we can create similar opportunities in core subjects — giving students the freedom to define authentic and purposeful goals for their learning, creating opportunities for students to lead that learning, and helping them to refine their work until it meets high standards of quality — the deeper their learning and engagement will be. (Mehta & Fine, 2019)

The evolution of high schools today can be seen as local school districts shift from being rote task masters to assisting students in securing their futures as competent employees or entrepreneurs. However, there is a struggle to break free from the traditional approach to education in order to do what is best for educating students. A barrage of assessments and the push to improve the scores of those assessments blocks students from developing creativity, critical thinking, and collaboration skills (Dintersmith & Wagner, 2015). Consequently, the path to success is vastly different from ten years ago, let alone 20 years ago. Dintersmith and Wagner (2015) note that, in the last century, adults needed to comprehend written information but seldom needed to verify that information; this is not the case nowadays, in a world where misinformation is commonplace. Now more than ever, students need to enter adulthood with deftness in analysis, critique, synthesis, and crafting a plausible and concise argument.

Therefore, the researchers chose to focus on the PSAT/NMSQT because this assessment has grown alongside the SAT in importance to the educational fabric of America, as both tests claim to identify students' college and career readiness. The power of the PSAT/NMSQT to propel students on different paths has increased over time. Naturally, with the multiple facets of the PSAT/NMSQT being applied to assess high school students' abilities, the mission of the current study is to determine the validity of the tool as a measurement of college readiness skills like problem formulation, knowledge acquisition and application, executive function and goal setting, and the ability to apply conceptual knowledge to real-world scenarios (Conley, 2010). Notably, after 2015, the SAT and the PSAT/NMSQT underwent a redesign to be more aligned to the expectations of 21st-century skills and career readiness. As stated on the College Board website, the redesigned PSAT/NMSQT and SAT purport to cover:

- Focus on the knowledge, skills, and understandings that research has identified as most important for college and career readiness and success
- Greater emphasis on the meaning of words in extended contexts and on how word choice shapes meaning, tone, and impact

This critically important redesign was completed in supposed concordance with the idea that "Work is no longer defined by your speciality; it's defined by the task or problem you and your team are trying to solve or the end goal you want to accomplish. Teams have to figure out the best way to get there - the solution is not prescribed" (Wagner, 2008). The foundation of the redesign to align with the modern-day expectations of 21st-century employees in the Information Age drives this research. Appropriately, the researchers would like to further examine the College Board's claims of assessing college-readiness skills of students using Hess' Cognitive

Rigor Matrix to clearly identify each PSAT/NMSQT question's skill assessment (Hess, 2014). The application of Hess' Cognitive Rigor Matrix to analyze the questions of the 2018 version of the PSAT/NMSQT will reveal the skills assessed by the tool and whether they accurately measure college-readiness skills based upon a cross-referenced table of Bloom's Taxonomy and Webb's Depth of Knowledge standards (Hess, 2014).

Research Question and Hypotheses

1. In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Reading Section questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?
2. In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Math section (both calculator and non-calculator) questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?
3. In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Writing section's questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?

Theoretical Framework

The structure of the study is predicated on various scholars' definitions and operationalization of college readiness skills through the lens of assessing cognitive rigor. The formation of cognitive rigor is based upon "a mainstay for over 50 years, Bloom's Taxonomy helps teachers formulate lessons that practice and develop thinking skills over a wide range of cognitive complexity" (Bloom, 1956). Although later revised by a team of education researchers headed by Anderson and Krathwohl (2001), the overall intent of the taxonomy remains:

categorize questions and activities according to their levels of abstraction. However, Bloom's Taxonomy suffers from limitations when selecting test items and formulating questioning strategies because it uses verbs to differentiate taxonomy levels — many verbs appear at multiple levels. They do not clearly articulate the intended complexity implied by the taxonomy. For example, in Bloom's original action verb taxonomy the word *select* appeared at the knowledge and comprehend level. A new model of rigor, depth of knowledge (DOK), fills this void by focusing on what the task requires of a student. The resulting combination of Bloom's Taxonomy and depth of knowledge — cognitive rigor — forms a comprehensive structure for defining rigor, thus posing a wide range of uses at all levels of curriculum development and delivery (Hess et al., 2009). The depth of knowledge that Hess refers to in her article directly connects to the higher-order thinking skills that the PSAT/NMSQT claims to assess by way of Webb's Depth of Knowledge Level 3 (Strategic Thinking) and Level 4 (Extended Thinking) are linked to Conley's structure of College and Career Readiness. Therefore, if the PSAT/NMSQT purports to be a tool that measures college and career readiness, the assessment should contain prompts that ask a student to reach Webb's Level 3 or Level 4 Depth of Knowledge. To further illustrate the link to the PSAT/NMSQT evaluation of higher-order thinking skills, the following except from *Now That's a Good Question! How to Promote Cognitive Rigor Through Classroom Questioning* by Francis (2016) beautifully demonstrates the similarities and differences between Bloom's Taxonomy and Webb's Depth of Knowledge:

> Cognitive rigor provides this enhanced educational experience by superimposing two academic frameworks that define how deeply students demonstrate their knowledge—Bloom's Revised Taxonomy and Webb's Depth-of-Knowledge. Bloom's taxonomy categorizes the *kind of knowledge* and *type of thinking* that students demonstrate to answer a question. Webb's Depth-of-Knowledge model designates the *depth of knowledge* that students express in a given context to answer a question. By aligning these two frameworks, cognitive rigor acts as a high-quality instructional tool to

> ensure teachers prepare their students for success in and out of the classroom. (Francis, 2016)

Francis (2016) also points out that cognitive rigor, used as a measurement tool of how intellectually involved students must be, can serve as a learning goal for students and their educators. The partnership of cognitive rigor and higher-order thinking skills can be displayed in Hess' Cognitive Rigor Matrix (HCRM) as "the categories in Bloom's taxonomy define the *subject matter* and describe the *skills* students must learn, whereas the levels of Webb's DOK model designate the scenario, setting, or situation in which students demonstrate and communicate their learning" (Francis, 2016). Consequently, the subject-specific HCRMs create a juxtaposition of a student working through the learning exercises' cognitive rigor and simultaneously applying a certain level of academic skill in order to complete the task. In essence:

> Hess (2013) describes the DOK levels as ceilings that designate how deeply students are expected to explain and use what they learn. Therefore, a higher DOK level does not necessarily mean it is "better" or even more desirable than other levels. It just provides a deeper context for the transfer and use of student learning. (Francis, 2016)

The researchers that authored this study believe that Francis's words bear repeating, "it is the job of educators to provide learning experiences that encourage such deep examination of knowledge. In this way, the use of cognitive rigor as a measurement" (Francis, 2016). Emphatically, cognitive rigor measures the higher-order thinking skills and depth of knowledge required to complete a task. Ergo, it is necessary to examine some of the most cogent explanations and applications of higher-order thinking skills and deeper learning to adequately design an assessment of the PSAT/NMSQT's validity. The next step to determining whether the PSAT/NMSQT questions are cognitively rigorous is to look into the definition of higher-order thinking skills through various lenses of research and how higher-order thinking skills are taught and operationalized in the classroom. Higher-order thinking skills have been studied, debated, and studied through the lens of psychology, Socratically contested in philosophy, and attempted to be defined, taught, practiced, and examined in education. In 1993, Lewis and Smith authored an article titled *Defined Higher Order Thinking,* attempting to rally all of the previous definitions of higher-order thinking skills and place them into a historical and categorical context. The authors describe cognitively-rigorous teaching as requiring students to transfer knowledge to new, open-ended, and complex problems, a skill necessary for success after graduating high school. However, Lewis and Smith (1993) and Dintersmith and Wagner (2015) note that, rather than providing cognitively-rigorous educations, students are instead asked to engage in lower-level thinking:

> When observing the way students are taught and evaluated, their school's unequivocal priority is to cover meticulously specified content. Classes and report cards are organized by subject. Tests revolve around content recall. Kids cram for days (or months, in the case of high stakes Standards of Learning, SATs, or APs) for tests that call for them to regurgitate content which they quickly forget. (Wagner & Dintersmith, 2015)

Invariably, the sole focus on assessment of content knowledge rather than the application of skills, in conjunction with a foundation of content knowledge, seems to hinder our students. Hess and colleagues (2014) state that there is little evidence regarding whether standards set forth by state-based and common core standards movements promote post-secondary success. They go on to posit that, while most agree that college and career readiness is crucial, few agree on what constitutes college and career readiness:

> Key weaknesses in the states-based standards movement were a major impetus for the [Common Core State Standards]: research showed that proficiency standards varied widely across states, while there was little evidence that the standards themselves fostered - or even were designed to foster - post-secondary success, which is considered

> necessary for individuals and the country to compete economically. College and career readiness has become the anchor concept for resolving these uncertainties about student achievement, and the CCSS are widely touted as that resolution.
> There is widespread agreement that focusing on college and career readiness means shifting from helping high school students become eligible for college (by taking the right courses, amassing credits towards graduation, and scoring well on the SAT) to getting them ready for college. Many questions remain, however, about what this really means. How, for instance, can we identify the critical skills and dispositions needed for deeper learning? And then how can we target and support their development? Once we know what students need, how can we help school systems and organizations assess the effectiveness of their efforts to prepare students for college and careers? (Hess, Gong, & Steinetz, 2014)

Consequently, the tool the researchers will use to assess the rigor of the PSAT/NMSQT combines the Revised Bloom's Taxonomy and Webb's Depth of Knowledge into one matrix. The result is a tool that captures the essence of both what a student is doing to expand their knowledge, and how they are applying their skills to perform a task. The Revised Bloom's Taxonomy uses "action words" that are thought to accurately describe the various cognitive processes that students and other thinkers use to learn and remember (Iowa State University, Center for Excellence in Learning and Teaching, 2020).

The idea of the student being able to apply their knowledge dynamically, rather than a passive reiteration of content, is the underlying concept of Webb's Depth of Knowledge:

> Webb designed his model as a means of increasing the cognitive complexity and demand of standardized assessments. Traditionally, standardized assessments measured students to think deeply about the academic content, concepts, ideas, and procedures they were learning. However, these assessments were limited in measuring students' ability to transfer and use what they were learning in different contexts. They were also limited in measuring the depth of understanding students must develop and demonstrate. Students were challenged to demonstrate – or show – the ability to think deeply about how to answer questions, address problems, accomplish tasks, and analyze texts and topics. However, they were not being challenged to communicate – or tell – how and why they could transfer and use what they were learning in different contexts. (Francis, 2017)

Definitions of Key Terms

Standardized Testing

- "A standardized test is any form of test that (1) requires all test-takers to answer the same questions, or a selection of questions from a common bank of questions, in the same way, and that (2) is scored in a "standard" or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. While different types of tests and assessments may be "standardized" in this way, the term is primarily associated with large-scale tests administered to large populations of students, such as a multiple-choice test given to all the eighth-grade public-school students in a particular state, for example.
- In addition to the familiar multiple-choice format, standardized tests can include true-false questions, short-answer questions, essay questions, or a mix of question types. While standardized tests were traditionally presented on paper and completed using pencils, and many still are, they are increasingly being administered on computers connected to online programs (for a related discussion, see computer-adaptive test). While standardized tests may come in various forms, multiple choice and true-false formats are widely used for large-scale testing situations because computers can score them quickly, consistently, and inexpensively. In contrast, open-ended essay questions need to be scored by humans using a common set of guidelines or rubrics to promote consistent evaluations from essay to essay—a less efficient and more time-intensive and costly option that is also considered to be more subjective" (The Glossary of Education Reform, 2014).

College Board

- The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity.
- Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education.
- Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT and the Advanced Placement Program. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools (College Board, 2020).

PSAT/NMSQT

- Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test
- A standardized test created by the College Board is used as a metric for students to practice skills and knowledge to gain an understanding of their preparation for the Scholastic Aptitude test that will be used for College Admissions.
- The PSAT is also used by guidance counselors and school administrators to place high school students into more rigorous courses, measure a high school's performance, and the PSAT is also used to grant students high school proficiency in becoming eligible for high school graduation.

SAT

- Scholastic Aptitude Test

- A standardized test created by the College Board that is used as a metric for ranking students in College admissions

AP

- Advanced Placement
- Advanced Placement courses are created by the College Board to provide High School students with a standardized experience in learning topics and skills that are equivalent to the first year of college course while still in High School. At the end of an Advanced Placement course, the students may elect to participate in an AP exam. If the students score high enough (a 5 is the highest score) on the AP exam, some colleges may accept that score as earning college credit for one college course (i.e., AP US History score of 4 or higher may be equal to one history course at a college).

ACT

- American College Testing
- Another standardized test that was created to compete with the SAT as a metric for ranking students in College admissions.

PISA

- Programme for International Student Assessment
- The PISA is a standardized test for high school age students that measures their ability to apply science, math, and reading skills to real world questions.

LSAT

- The Law School Admission Test (LSAT) is an integral part of law school admission in the United States, Canada, and a growing number of other countries. The purpose of the

LSAT is to test the skills necessary for success in the first year of law school. Those skills include reading comprehension, reasoning, and writing, and the test results help admission decision-makers and candidates alike gain valuable insight into law school readiness (Law School Admission Council, 2020).

- The LSAT is administered in two parts. The first part of the test is a multiple-choice exam that includes reading comprehension, analytical reasoning, and logical reasoning questions.

HCRM

- Hess' Cognitive Rigor Matrix
- Hess' Cognitive Rigor Matrix assists teachers in applying what cognitive demand might look like in the classroom and guides test developers in designing and aligning test items and performance tasks. Content-specific descriptors in each of the CRMs are used to categorize and plan for various levels of abstraction – meaning an analysis of the mental processing required of assessment questions and learning tasks (Hess, 2014).

DOK

- "Webb's Depth of Knowledge categorizes tasks according to the complexity of thinking required to complete them successfully. Recently, educators have begun applying Webb's DoK to help them design better instruction.
    - Level 1. Recall and Reproduction: Tasks at this level require recall of facts or rote application of simple procedures. The task does not require any cognitive effort beyond remembering the right response or formula. Copying, computing, defining, and recognizing are typical Level 1 tasks.

- Level 2. Skills and Concepts: At this level, a student must make some decisions about their approach. Tasks with more than one mental step, such as comparing, organizing, summarizing, predicting, and estimating, are usually Level 2.
- Level 3. Strategic Thinking: At this level of complexity, students must use planning and evidence, and thinking is more abstract. A task with multiple valid responses, where students must justify their choices, would be Level 3. Examples include solving non-routine problems, designing an experiment, or analyzing characteristics of a genre.
- Level 4. Extended Thinking: Level 4 tasks require the most complex cognitive effort. Students synthesize information from multiple sources, often over an extended period of time, or transfer knowledge from one domain to solve problems in another. Designing a survey and interpreting the results, analyzing multiple texts to extract themes, or writing an original myth in an ancient style would all be examples of Level 4" (Aungst, 2014).

Bloom's Taxonomy

- Bloom's Taxonomy is a classification system used to define and distinguish different levels of human cognition—i.e., thinking, learning, and understanding. Educators have typically used Bloom's taxonomy to inform or guide the development of assessments (tests and other evaluations of student learning), curriculum (units, lessons, projects, and other learning activities), and instructional methods such as questioning strategies (The Glossary of Education Reform, 2014).

Cognitive Rigor

- “Cognitive rigor measures the depth and breadth of topics that should be taught as part of an educational experience based on the following criteria:
    - Complexity of the concepts and content taught and obtained
    - Kind of knowledge acquired
    - Type of thinking demonstrated
    - Depth of knowledge communicated” (Francis, 2016)

Higher-Order Thinking

- “Higher-order thinking skills go beyond basic observation of facts and memorization. They are what we are talking about when we want our students to be evaluative, creative, and innovative. When most people think of critical thinking, they think that their words (or the words of others) are supposed to get “criticized” and torn apart in argument, when in fact all it means is that they are criteria-based. These criteria require that we distinguish fact, from fiction; synthesize and evaluate information; and clearly communicate, solve problems and discover truths.
- Using Bloom’s Taxonomy of thinking skills, the goal is to move students from lower- to higher-order thinking:
    - from knowledge (information gathering) to comprehension (confirming)
    - from application (making use of knowledge) to analysis (taking information apart)
    - from evaluation (judging the outcome) to synthesis (putting information together) and creative generation.
- This provides students with the skills and motivation to become innovative producers of goods, services, and ideas. This does not have to be a linear process but can move back

and forth, and skip step" (University of Connecticut Center for Excellence in Teaching and Learning, 2020).

CCR

- College and Career Readiness
- "College and career readiness can be defined as success –without remediation–in credit-bearing general education courses or a two-year certificate program. "Succeed" is defined as being able to progress successfully in the chosen program.
- Conley created four Keys of College Readiness
    - Key Cognitive Strategies
        - Problem formulation, research, interpretation, communication, precision and accuracy
    - Key Content Knowledge
        - Key terms & terminology, factual information, linking ideas, organizing concepts
    - Key Learning Skills & Techniques
        - Time management, study skills, goal setting, self-awareness, persistence, collaborative learning, student ownership of learning, technological proficiency, retention of factual information
    - Key Transition Knowledge & Skills
        - Postsecondary program selection, admissions requirements, financial aid, career pathways, postsecondary culture, role & identity issues, agency" (Conley, 2011)

Nature of the Study

An investigation into the effective application of Hess' Cognitive Rigor Matrix (HCRM), Bloom's Taxonomy, and Webb's Depth of Knowledge (DOK) is a prerequisite not only for the researcher to perform the study with proficiency but for also the accuracy of the results of the study. Subsequently, a seasoned assessment evaluator with a strong background in the application of HCRM will assist in determining the inter-rater reliability of the PSAT/NMSQT evaluation. Dr. Michael Thumm has graciously offered his support as the double reader for the PSAT/NMSQT evaluation. The researchers made the conscious choice to utilize HCRM for the conceptual framework as the investigation of the PSAT/NMSQT takes place. Conveniently, Webb's Depth of Knowledge and Bloom's Taxonomy have both attempted to measure the number of critical thinking skills a student is asked to use to complete a task or solve a problem. However, "Bloom's thinking levels and Webb's Depth of Knowledge differed in scope, application, and possibly intent" (Hess, 2018). Subsequently, HCRM is a summation of the following points:

- Bloom's Taxonomy categorizes the cognitive skills required of the brain to perform a task, describing the 'type of thinking processes' necessary to answer a question or complete a task.
- The types of thinking identified in Bloom's Taxonomy CAN lead to deeper learning when matched with increasingly more complex content: deeper understanding, deeper application, or deeper analysis.
- Depth of Knowledge focuses on how deeply students need to know and interact with content to generate a specific type of response. DOK levels relate more closely to the

depth of content understanding and scope of a learning activity, which manifests in the skills required to complete a task from inception to end.

- DOK levels are not sequential or linear. Students need not fully master content with Level 1 tasks before doing Level 2 tasks. In fact, giving students an intriguing Level 3 or Level 4 task prompt can provide context and motivation for engaging in the more routine learning at DOK Levels 1 and 2 needed to be successful with the more complex task (Hess, 2018)

The HCRM model creates a connection between Webb's Depth of Knowledge and Bloom's Taxonomy. The different levels identified in each are merged to give a more well-rounded focus for educators when providing higher-order thinking and problem-solving activities and questions for students. To facilitate an evaluation of the PSAT/NMSQT through the lens of the HCRM, the researchers anticipate a more thorough analysis of the college readiness skill requirements placed on the students to complete the prompt or question successfully.

> Furthermore, a related study had worked diligently to highlight an ongoing effort at the College Board to establish college readiness benchmarks on the SAT[R], PSAT/NMSQT[R], and ReadiStep as well as to provide schools, districts, and states with a view of their students' college readiness. College readiness benchmarks were established based on SAT performance, using a sample of approximately 68,000 students across 110 four-year institutions. The college readiness benchmark was calculated as the SAT score associated with a 65 percent probability of earning a first-year Grade Point Average of 2.67 (B-) or higher. (Wyatt et al., 2011)

The study conducted by Wyatt and colleagues (2011) is one of many quantitative studies that correlate standardized testing to college GPA. The researchers chose to examine the fundamental skills students need to be successful in the 21st-century workplace. Recently, there has been a shift in education in which the student-centered, inquiry-based, open-ended, and problem-solving classroom has become the basis for pedagogy to enhance the student experience

(Hess, Colby, & Joseph, 2020). Thus, the gap in educational research exists where the actual skills students need to be successful on a standardized test are effectively measured using Hess' Cognitive Rigor Matrix (HCRM).

A qualitative design was the appropriate method for this study as the researchers sought to develop a deeper understanding of the rigor of the PSAT/NMSQT using the cognitive rigor matrix. Moreover, a document analysis approach was chosen to support the qualitative research involved in this investigation of the 2018 PSAT/NMSQT. The study's rationale falls into the sector of Evaluation Research, as the researchers in this study partnered to investigate an assessment's effectiveness at measuring college-readiness skills. The PSAT/NMSQT has received increased notoriety due to its long-term implications for students, community stakeholders, and school rankings. Additionally, the literature review indicates a lack of studies extensively evaluating the SAT and ACT. Yet, the PSAT/NMSQT has been studied less frequently in terms of skills assessed by the examination. Most importantly, the PSAT/NMSQT provides a score report which indicates the difficulty of a question. This particular study may not determine, quantitatively, the direction and strength of the relationship between the PSAT/NMSQT questions and cognitive rigor, but may provide other researchers with a springboard to conduct additional studies on standardized tests' validity. The study focuses on the quality of the College Board's PSAT/NMSQT measurement tool of college readiness using HCRM. Therefore, the coding scheme will use the x-axis and y-axis of the Rigor Matrix. The x-axis (Webb's DOK score) and y-axis (Bloom's Taxonomy score) will be discussed with the other researcher and sorted on a spreadsheet. When data analysis is complete, the results will then be collated in a spreadsheet, reviewed through the research lens, and summarized to reveal the effectiveness of the PSAT/NMSQT's measurement of college-readiness skills.

Limitations

A limited number of studies assessed the higher-order thinking skills required of students to perform well on standardized tests. There were no studies that applied a tool like the HCRM to the PSAT/NMSQT. The vast majority of the research examined correlations between SAT scores and environmental factors or used SAT scores to predict student outcomes. One of the limitations of the current study is that only one year of the PSAT/NMSQT will be studied. The sample of questions is composed of every single question of the 2018 PSAT/NMSQT. The majority of the questions were multiple choice. The very nature of multiple choice, with the answer being available to a student to recognize or select based on familiarity, may limit the ability of the test-taker to access higher-order thinking skills. Additionally, this study only reviewed the PSAT/NMSQT, and there are several other high school standardized assessments that intend to evaluate students' college and career readiness. This particular test was chosen for the far-reaching impact it can have on high school sophomores and juniors in terms of meeting proficiency needed to graduate high school graduation higher-level courses, as well as the evaluation of school districts in the state of New Jersey. Also, this College Board product specifically markets itself as a tool that identifies and assesses college and career readiness for the modern-day student. Lastly, this study was developed during 2018 and 2019, and the study contains only the prompts and questions from the 2018 PSAT/NMSQT.

Ethical Considerations

The researchers worked diligently to provide a balanced perspective on the advantages and disadvantages of standardized testing of high school students. The literature review contains studies from professionals in the educational research field that explain both the positives of standardized testing and the downsides of attempting to measure a high school student's skills and knowledge using standardized tests. Significantly, the foundation of the study is based upon Hess's Cognitive Rigor Matrix (HCRM), which measures the cognitive rigor of each question on the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT) as a validated and reliable tool that has withstood the test of time. Notably, in two large-scale studies, the HCRM has also been found to be an unbiased measure, which will lead to accurate validation of the PSAT/NMSQT (The Standards Company LLC, 2008a, 2008b). Thus, the ethical integrity of the study is solidified through the training on and application of the HCRM. Regarding the training on and application of the HCRM, the researchers will participate in multiple training sessions to hone their skills in applying the HCRM to multiple-choice and open-ended standardized questions. Lastly, as this is a document study, there are no participants in the study other than a standardized test that has been made public since December 2018.

Significance of the Study

The significance of this study will aim to shed light on the standardized testing system that runs concurrently with local, state, and national education policy and practice. The term 21st-century skills have been applied to multiple facets of education. In the state of New Jersey, the final report of the New Jersey Department of Education College and Career Readiness Task Force defined career and college readiness as "The knowledge and skills that high school graduates must possess in English and mathematics— including, but not limited to, reading, writing, communications, teamwork, critical thinking, and problem-solving to be successful in any future endeavors" (NJDOE, 2012). The PSAT/NMSQT website breaks career and college readiness into three sections (*Figure 2*):

Table 2. Details of the three PSAT/NMSQT sections adapted from www.Collegeboard.org (2018).

Reading Test	The Reading Test focuses on the skills and knowledge at the heart of education: the stuff you've been learning in high school, the stuff you'll need to succeed in college. It's about how you take in, think about, and use information.
Writing Test	When you take the Writing and Language Test, you'll do three things that people do all the time when they write and edit: Read. Find mistakes and weaknesses. Fix them. The good news: You do these things every time you proofread your own schoolwork or workshop essays with a friend. It's the practical skills you use to spot and correct problems—the stuff you've been learning in high school and the stuff you'll need to succeed in college—that the test measures.
Math Test	Instead of testing you on every math topic there is, the PSAT/NMSQT/NMSQT and PSAT/NMSQT 10 ask you to use the math that you'll rely on most in all sorts of situations. Questions on the Math Test are designed to mirror the problem solving and modeling you'll do in: ● College math, science, and social science courses ● The jobs that you hold ● Your personal life For instance, to answer some questions you'll need to use several steps—because in the real world a single calculation is rarely enough to get the job done.

The power of the PSAT/NMSQT test has been expanded when considering graduation requirements, access to Advanced Placement class, access to rigorous course loads, and the attempt to equip New Jersey high school students with college and career readiness skills within the ever-evolving curriculum. Teachers, administrators, students, and parents/guardians are now encountering different approaches to embed college readiness skills into the everyday curriculum as a result of the evolving 21st-century workplace shift into a digital age. Moreover, the private sector has taken full advantage of the pressure for students to perform on standardized tests and has reaped the benefits of millions of dollars. The current research shows a gap in the identification of how standardized assessments ask students to apply their college-readiness skills. Therefore, the absence of academic studies on this particular facet of standardized testing, specifically the PSAT/NMSQT, leaves school districts and community stakeholders in a place where budgetary funding decisions, curricular decisions, professional development investment, and pedagogical strategies are based on a standardized test that has not been vetted through the lens of well-validated, theoretically-driven, and empirical tools designed to measure cognitive rigor, such the HCRM.

To understand higher-order thinking skills and the application of those skills to college and career readiness, one may find it beneficial to understand deeper learning:

> Deeper learning is an umbrella term that has emerged over the past decade to encompass a range of desirable attributes of schooling, attributes rooted in the premise that schooling needs to move beyond rote learning and shallow testing. The Hewlett Foundation, which helped to popularize the term, defines it as those combined characteristics of schooling that enable learners to 'develop significant understanding of core academic content, exhibit critical thinking and problem-solving skills, collaborate, communicate, direct their own learning and possess an academic mindset'" (Mehta & Fine, 2019).

By the same token, Wagner (2008) defines career and college readiness skills with a paradigm of "The New World of Work and the Seven Survival Skills":

1. Critical Thinking and Problem Solving
2. Collaboration Across Networks and Leading by Influence
3. Agility and Adaptability
4. Initiative and Entrepreneurialism
5. Effective Oral and Written Communication
6. Accessing and Analyzing Information
7. Curiosity and Imagination

To emphasize and synthesize the deeper learning in concordance with the viewpoints of Mehta and Fine (2019) and Wagner to a slightly more simplistic, direct, and operational definition, the researchers found Brookhart's phraseology to be effective in detailing higher-order thinking skills. In her book *How to Assess Higher Order Thinking Skills in Your Classroom,* Brookhart states, "students can apply the knowledge [and skills] they developed during their learning to new contexts" (Brookhart, 2010). Naturally, Brookhart's definition is the straightest line between two points but slightly misses the nuance of the cognitive rigor that complements higher-order thinking skills in which students activate the thought process and the depth of knowledge to consciously transfer knowledge and skills to solve a problem. As a result, the student expands their higher-order thinking skills both in the depth of their knowledge application and in the breadth of skills applied to solve the problem - this is the core of the research study using the HCRM to analyze PSAT/NMSQT.

Accordingly, the HCRM was used to assess the cognitive complexity of the questions provided in an online-based program used in a similar study. This related study's model was used because the HCRM referred to both Bloom's Revised Taxonomy and Webb's Depth of Knowledge, two renowned educational frameworks that evaluate student skills, and provides a more detailed description of keywords and tasks that could be asked for each level of cognitive complexity. The HCRM offers a more thorough analysis of each question assessed from the online-based program, resulting in a better understanding of the types and levels of cognitive complexity provided to students (Sydoruk, 2018). Subsequently, the researchers aimed to analyze the cognitive complexity of 113 questions and open-ended prompts presented to students through the PSAT/NMSQT that aim to evaluate a student's career and college readiness. Again, the focus of this qualitative document analysis study was structured intentionally to identify the depth of the college and career readiness skills within the PSAT/NMSQT. This study will give researchers more data on the PSAT/NMSQT's ability to assess higher-order thinking skills for high school students, as well as a foundation for research to structure assessments that compel students to access higher-order thinking skills in a standardized testing environment.

Summary

Chapter I provided insight into the PSAT/NMSQT's ability to measure students' college readiness skills and the application of the HCRM to measure the validity of the claim made by the College Board that the PSAT/NMSQT is a rigorous assessment of higher-order thinking skills. Chapter II contains a literature review on various viewpoints regarding higher-order thinking skills, how to measure higher-order thinking skills, theories regarding student outcomes and the skills necessary to propel students to success beyond high school graduation, other studies on the PSAT/NMSQT, the advantages and disadvantages of standardized testing related to college and career readiness, and how higher-order thinking skills are taught in schools today. Chapter III details the design methodology of the study and begins with a review of the details provided in Chapter I. The PSAT/NMSQT of 2018 will provide the data for the study, and the data will be analyzed using the HCRM. Chapter IV presents the data and statistics of the study. Chapter V synthesizes the outcomes of using the HCRM to evaluate the higher-order thinking skills necessary to complete a PSAT/NMSQT question and further reviews the results to identify the validity of the PSAT/NMSQT as a measure for college and career readiness. Finally, the dissertation concludes with suggestions for further studies and recommendations that could advance the study of measuring higher-order thinking skills.

Chapter II

LITERATURE REVIEW

Introduction

The literature review had three main purposes. First, the literature review serves to illustrate the evolution of education's focus on higher-order thinking skills as the world shifts from a computation-based economy to an innovation-based economy. Second, an investigation into the PSAT/NMSQT assessment tools, its design, and the study of the PSAT/NMSQT was completed to show the growing significance and implications of assessing a student's higher-order thinking skills using the PSAT/NMSQT. Thirdly, the literature review provides detailed applications of how Hess' Cognitive Rigor Matrix uses components of Bloom's Taxonomy and Webb's Depth of Knowledge. Therefore, a review of the literature regarding the application of Hess' Cognitive Rigor Matrix was required to show the context in which the tool has functionality as, in the current study, the researchers analyzed the strength of the PSAT/NMSQT's questions for cognitive rigor by using Hess' Cognitive Rigor Matrix.

Literature Search Procedures

For the current study, we used procedures for scholarly literature reviews that involves searching for articles on online databases using keywords such as PSAT/NMSQT, college readiness, assessing higher-order thinking and critical thinking, Webb's Depth of Knowledge, Bloom's Taxonomy, and Hess' Cognitive Rigor Matrix. This approach is in keeping with approaches recommended by others who have researched this topic (Sydoruk, 2018).

Criteria for Inclusion of Literature

Again, the researchers have based their dissertation on the structure provided by Sydoruk's (2018) study:

- Peer-reviewed research, including dissertations
- Theoretical literature
- Publications intended for educators, such as textbooks and mainstream professional development books
- Empirical studies that focused on higher-order thinking skills in regards to teaching practices or student activities
- Studies that focused on K–12 public school districts and their students
- Quantitative and qualitative studies, including case studies and comparative studies
- Studies conducted after 2000
- Works published by theorists regarding taxonomies, frameworks, and/or models for higher-order thinking

Review of Literature Topics

According to Wagner (2008):

> The way work is organized now is lots of networks of cross-functional teams that work together on specific projects. Work is no longer defined by your speciality; it's defined by the task or problem you and your team are trying to solve or the end goal you want to accomplish. Teams have to figure out the best way to get there-the solution is not prescribed. (Wagner, 2008)

There has been a shift from being a competent employee or employer towards being an agile visionary who excels at communication and collaboration. Nevertheless, a current debate can be found regarding which higher-order thinking skills are required to be successful in the 21st-century economy. Monolithic institutions of our culture such as newspapers, taxi companies, big-box appliance stores, retail stores, cable television, and shopping malls are becoming less and less prevalent due to the influx of disruptors who can seek, analyze, and act upon data more efficiently. Consequently, the path to success is vastly different than it was ten years ago, let alone 20 years ago. College-readiness, higher-order thinking skills, and critical thinking skills, as well as the tools to measure the advanced skills of students, have been frayed in several directions, "For too long, students have been passive consumers and recipients of whatever adults give them: books, facilities, knowledge, tests, and disciplines. Schools have been built to facilitate effective consumption and create great recipients rather than makers, creators, and entrepreneurs" (Zhao, 2012). The disconnect between how to teach and check for understanding of knowledge can be seen in Susan M. Brookhart's book, *How to Assess Higher-Order Thinking Skills in Your Classroom* (2010): "definitions that I find helpful fall 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" (Brookhart, 2010). A recent New York Times article underscores the burden of bored, disengaged, bewildered, and disaffected students at high schools across the U.S., in which the article echoed the sentiment stated by Brookhart in terms of the dire need for students being able to engage in open-ended, real-world problems. Mehta and Fine (2019) similarly noted that, in particular, students do not get the opportunity to independently lead their own exploration of academic or extracurricular material and thus students are inhibited from engaging in deep learning.

The College Board's reaction to the new wave of education and economics was to redesign their assessment. Designers of the SAT, PSAT/NMSQT, and AP Testing, the College Board play a strong role in a high school student's post-graduation planning and pathways. Notably, the main landing webpage of the PSAT/NMSQT on www.collegeboard.org has a subpage titled, Inside the Test, which claims the following:

> The PSAT/NMSQT and PSAT/NMSQT 10 are highly relevant to your future success because they focus on the skills and knowledge at the heart of education. They'll measure what you learn in high school and what you need to succeed in college. If you think the key to a high score is memorizing words and facts you'll never use in the real world, think again. You don't have to discover secret tricks or cram the night before. The best way to prepare for the test is to take challenging courses, do your homework, prepare for tests and quizzes, ask and answer lots of questions. In short, take charge of your education and learn as much as you can. (College Board, 2018)

In essence, the PSAT/NMSQT claims to be a tool that is not only integral in the future of a high school student's success but also can evaluate a student's array of higher-order thinking skills.

Significantly, the challenge to any standardized assessment is to remain at the cutting edge of education in which students are asked to demonstrate the ability to transfer their knowledge and skills to solve complex, open-ended problems. The excerpts detailing the PSAT/NMSQT's relevance to future success indicate a lack of faith in standardized tests' ability to measure college readiness skills successfully. However, when one digs deeper into the College Board website, the redesigned SAT claims to engage students in close reading, and honor the best work of the classroom by asking students to perform evidence-based reading, and writing as the students are required to interpret, synthesize, and use evidence found across a wide range of sources. Furthermore, on the redesigned SAT, the students are asked to find the best answer to a question by collecting information conveyed in words and graphics. The Writing and

Language section challenges students to analyze a series of sentences or paragraphs and decide if it makes sense or if they should edit a part of the accompanying passage so that it clearly and accurately communicates the information in the graphics. Interestingly, the SAT has refined the Problem Solving and Data Analysis portion of the exam to measure how quantitatively literate a student has become over the course of their educational career. Essentially, the student is asked to decipher ratios and percentages and to use proportional reasoning to solve problems in science, social science, and career contexts. Lastly, an upgraded Math section features multistep applications to solve problems in science, social science, career scenarios, and other real-life situations. The test sets up a scenario and asks several questions that give the student an opportunity to dig in and model a real-world problem mathematically (College Board, 2018).

Concurrently, the evaluation of high schools in the United States is in part based upon standardized testing statistics. Specifically, the New Jersey Performance Report Rationale details the nature of their evaluation as part of New Jersey's accountability system. Essentially, the NJDOE annually publishes a large amount of data regarding school performance reports to present a comprehensive picture of overall school performance, and the College and Career Readiness section provides information about how student behaviors correlate with greater success in college and career (NJDOE, 2018). Subsequently, the SAT, ACT, and the PSAT/NMSQT are the main tools used to gather data regarding college and career readiness for the NJDOE. Additionally, *New Jersey Monthly Magazine* is a publication that constructs a metric every two years in which schools are arranged based on test scores and college placement. Lastly, the New Jersey Department of Education (2016) has stated that the State Board of

Education will utilize the PSAT/NMSQT as one of the metrics for making students eligible for high school graduation. Naturally, with multiple organizations utilizing the data of the PSAT/NMSQT to measure high school students' abilities, the mission is to review the validity of the tool as a measurement of college readiness skills.

However, in stark contrast to the claims of the College Board's website, researchers have found that content knowledge, now readily available through the internet, is no longer valued in the world place. Instead, the ability to make connections, apply concepts, efficiently design, and creatively innovate new ideas or products based upon content knowledge is the high-valued currency in the workplace (Wagner & Dintersmith, 2015).

Significantly, the authors of *Most Likely Succeed* (2015) go on to claim:

> The standardized testing movement is failing American students, to test millions of students every year is expensive in terms of time, money, and opportunity cost. With the goal of rank-ordering millions of test-takers, assessment inevitably gets reduced to simple multiple choice quizzes, even though there's a complete disconnect between what is easy to test and what really matters. (Wagner & Dintersmith, 2015)

One can see the juxtaposition of the College Board claims to be an accurate tool of student skills and the insinuation of Dintersmith that the standardized testing model hinders student growth and ill prepares youth with a chance to grow as open-ended, complex problem solvers.

Theoretical Framework

The design of the literature review is to illustrate and detail the fundamentals of the skills required for high school students to enjoy future prosperity and study different assessment tools in their validity in examining student skills. Through the lens of history, questions abounded in the field of education on how to push students to become positive contributors to an ever-changing economy and forthright citizens in a seemingly polarized, democratic society:

While the term 'deeper learning' is new, many of the aspirations it represents are

> longstanding. For instance, Paulo Freire, in 1970, decried the tendency of 'banking' models of pedagogy, where children are treated as empty vessels in which teachers 'deposit' knowledge, and argued for 'problem-posing' as an alternative. Alfred North Whitehead, in 1929, discussed the difference between 'active' forms of learning and 'inert' knowledge. Joseph Mayer Rice, in 1893, contrasted 'old education,' which emphasized drilling and recitation, with 'new education,' which aimed 'to lead the child to observe, to reason, and to acquire manual dexterity as well as to memorize facts-in a word, to develop the child naturally in all his faculties. Modern scholars describe this contrast between 'ambitious instruction,' which asks students to reason and understand underlying conceptual structures, and 'conventional instruction,' which does not. While there are some differences among these formulations, in a fundamental way they share an emphasis on 'deep' versus 'shallow' education, that is on education that asks students to think versus education that asks them to follow directions, and education that has purpose and meaning for students versus education that does not. (Mehta & Fine, 2019)

Mehta and Fine (2019) provide a conceptual framework in which students are asked to grapple with, manipulate, collaborate, lean into, and lunge at open-ended, complex problems in order to achieve a deeper level of learning. The conventional approach to education places students in a deficit through the lens of innovation, communication, critical thinking, and creativity. The students in today's world should be equipped with skills to apply their tools to new situations that may not have a single correct answer.

Why the emphasis on deeper learning? What makes the need for a dramatic shift so pertinent? What tectonic forces have compelled a century of instructional practices to be necessarily altered? According to Mehta and Fine (2015), economic changes have decreased many traditionally middle-class jobs, meaning that students now need post-secondary education merely to be competitive with others on the job market. They also note that the skills that today's employers value are different from the skills that were valued in the past and that equity-related issues have increased focus on trying to boost the performance of students (especially low socio-economic status students and students of color) within the public school system (Mehta &

Fine, 2019). Finally, Mehta and Fine (2019) note that the complex, 21st-century issues that today's students will need to face will require higher-order thinking skills such as creativity, communication, collaboration, adaptability, and empathy.

In addition, Daniel Pink's book, *A Whole New Mind: Why Right-Brainers Will Rule the Future* (2005), opens with an argument that states:

> The last few decades have belonged to a certain kind of person with a certain kind of mind-computer programmers who could crank code, lawyers who could craft contracts, Masters of Business Administration (MBAs) who could crunch numbers. But the keys to the kingdom are changing hands. The future belongs to a very different kind of person with a very different kind of mind-creators and empathizers, pattern recognizers, and meaning makers. These people-artists, inventors, designers, storytellers, caregivers, consolers, big picture thinkers-will now reap society's richest rewards and share its greatest joys. We are moving from an economy and a society built on the logical, linear, computerlike capabilities of the Information Age to an economy and society built on the inventive, empathetic, big-picture capabilities of what's rising in its place, the Conceptual Age. (Pink, 2005)

Again, what is necessitating such a radical shift? What about today's educational systems and typed approaches of pedagogy have been deemed insufficient by prospective employers. The idea, boiled down to its most common denominator, is that American society has been structured for factory-like standardization and role-defined boxes for employees. The work that had been successful was based upon a currency of content knowledge. The multiple outlets of information and the handheld nature of technology no longer demand a currency for content, but the world has shifted to communication, collaboration, empathy, creativity, synthesis, and adaptability being the kings of currency in the latest version of the global economy (Pink, 2005).

To that end, the researchers circle back to the research question: does the PSAT/NMSQT assess higher-order thinking skills through the lens of cognitive rigor it purports to assess? Invariably, this study has been propelled forward because the vast majority of the research completed on the PSAT/NMSQT seeks to quantitatively link PSAT/NMSQT score performance and college grade point averages or to quantitatively link between the PSAT/NMSQT score and the student's expectancy to successfully complete College Board's Advanced Placement exams. In a study that expanded upon the work of Wiley, Wyatt, and Camara (2010) created a predictive model called the Academic Readiness Indicator (ARI) that combined high school grade point average (HSGPA) and SAT Scores. The college-ready benchmark was placed at a B+ HSGPA, a composite score of 1550 on the SAT (Proctor et al., 2009). Invariably, the college-readiness measurement tool is only using the score the students earned on the PSAT/NMSQT and the student's GPA, not an analysis of the actual questions on the examination in their pursuit of the definition of "college-ready." Conclusively, research on the test questions contained on the PSAT/NMSQT has been quite barren.

Before the literature review pivots to the details of high stakes, standardized tests (both the benefits and the limitations), the definition and operationalization of cognitive rigor, and different tools used to assess cognitive rigor must be highlighted. The description of educational practices found in Sizer's *Horace's Compromise: The Dilemma of American High Schools,* written in 1984, perfectly details the concerns faced by today's educators even though decades have passed since the book was published: the idea that teachers continue to explain things to students, rather than coach students and encourage them to independently transfer skills and knowledge to solve open-ended, complex problems.

Sizer (1984) details the discordance of student learning and the way students are assessed in schools:

> An English program that assigns only short stories would give students little scope either in learning to observe and analyze a variety of means of written expression or in exploring models for them to use in their own writing. A science course built on sheer

> memory of work that never gives examples of or experiences of scientific inquiry would be as stunted as a course that engages in some sort of disembodied, abstract problem-solving that demands of the students no command of precise knowledge.

Sizer (1984) goes on to explain not only should education sharpen students' skills to be accessed across disciplines and multi-level learning experiences, but also:

> The subject matter chosen should lead somewhere, in the eyes and mind of the student. This means that it must connect to wherever that student is rooted-his experience-and that it promises to take him toward an important place. It must be ultimately useful and patently interesting to him, at the time it is learned in the future...all subject matter should be fuel for the travel to more subject matter. The details of every school curriculum must meet this test. There are few universal answers, though. How subject matter is generated depends on where a particular student or group of students is. The construction of the subject matter of any curriculum is a task of cabinetmaking, not of prefab carpentry. The pieces *have* to fit the conditions of a peculiar population to each school. Master plans for cities, states, and the nation that standardize instruction are certain to be inefficient: no one set of procedures can conceivably serve most students well. (Sizer, 1984)

Limits of High Stakes Standardized Testing

The use of standardized testing in the U.S. dates back to the U.S. Army's development of the Alpha Test, which was used to screen officer candidates during World War I; from there the practice moved into classroom settings and has continued during the one hundred years since (Tienken & Orlich, 2013). Today, standardized testing is accepted almost innately as a quantifiable and qualitative measure of a student. The standardized test is used as a photo album. Each test score is a snapshot of a student's strengths and weaknesses. In today's educational world, national and state-level legislation has locked the standardized test into existence. Inherently, the high-stakes testing of the education system in the United States of America seems to not structure time for students, teachers, administrators, parents, or community stakeholders to have meaningful conversations about the test results in a constructive manner that promotes growth. As a result, the snapshot taken by one test score is a frozen moment in time that can either propel a student to new opportunities or hold a student down, or even worse, keep a middling student stagnant. The evaluation of a school's annual yearly progress is largely based upon the standardized test scores of the students in the school. Tienken (2008) found a strong correlation not in student ability or pedagogy, but rather the number of students who are eligible for free and reduced lunch. Essentially, the families' socioeconomic status in a school's district plays a large role in determining the outcomes of standardized test scores (Tienken, 2008). Simply based on one's socioeconomic status, the scores of those children differed on a standardized test. Therefore, the research on the validity of the questions on a standardized test like the PSAT/NMSQT becomes that much more amplified.

New types of assessment are required to measure new content and skill-based frameworks. Still, the cost-effective solution tends to be multiple-choice, "bubble sheet" exams rather than the more labor-intensive essay responses. Jean Anyon's classic text "Social Class and School Knowledge" describes four types of schools—working class, middle class, affluent professional, and executive elite—that nearly parallel both the depth of knowledge and the rigor and relevance frameworks in an unsettling metaphor. At the working class/DOK 1/Quadrant A rigor and relevance end of the spectrum, students are expected to acquire information and offer basic regurgitations of generalizations that support applications in assembly line or manual labor settings. Contrastingly, at the other end of the spectrum, executive elite/DOK 4/Quadrant D rigor and relevance end, students are expected to provide multivariate analysis and reasoning toward the planning and execution of solutions to complex problems. One can clearly see the disparity.

Suppose multiple-choice tests are the best measure of student achievement. Why is it that students in working-class schools, who receive daily doses of the type of decontextualized, rote knowledge bits of information most readily measured by high-stakes tests, not doing as well on these tests as students from more economically-privileged schools? Why do they not "achieve" at higher rates than students at affluent or executive elite schools? Short of the "pollyannaish view of the potential for [perfect] coherence and alignment of state policy" that Grant and Salinas critique, educational stakeholders should demand that curriculum, teacher professional development, teacher certification, and student assessment be mutually reinforcing toward greater depths of student knowledge for every child in every school setting (View, 2012).

Astoundingly, Henderson (2005) researched the Law School Admission Test (LSAT) and discovered some eye-popping anomalies. "The LSAT is a cultural lightning rod. While some prominent scholars attack the test as a poor predictor of law school success that is biased in favor of the privileged, others praise it as a valuable tool for social mobility" (Henderson, 2005). Yet another example of a high-stakes test that takes a single snapshot of one's ability in a constrained amount of time and places the test-taker on the knife edge to propel their future or create an obstacle to their occupational and socio-economic goals. Essentially, "based on a detailed sample set drawn from a national and a regional school, the [research by William Henderson] presents strong empirical evidence that test-taking speed, perhaps more so than what the LSAT purports to measure (i.e., reasoning skills), affects student performance on both the LSAT and actual law school exams (Henderson, 2005). Ironically, the swift nature of the LSAT conflicts with the deliberate, methodical, and meticulous research an effective law practice entails. Subsequently, Henderson (2005) provides another example of how high-stakes, standardized testing is subject to variables beyond higher-level thinking, such as the ability to work quickly under pressure; these factors may or may not be important to a students' future career goals.

More evidence of the disconnect between standardized testing outcomes and the test-takers' ability resides in the realm of special education. Students with special needs are tested under strained environmental conditions that are not in line with their Individualized Educational Plan (IEP). Furthermore, having an IEP does not necessarily guarantee you receive accommodations on a College Board-based test. The College Board has its own internal review of a student's IEP to provide the College Board's version of accommodations for students with special needs. Conclusively, demographics, behavior, critical thinking skills, and IEP status affect both PSAT/NMSQT scores and GPA. Both non-academic factors, particularly critical thinking, alongside academic indicators, should be used to measure and evaluate Career and College Readiness (CCR). One research study suggests that for high school students without disabilities, there is a link between critical thinking and traditional academic indicators of CCR. For these students, the overall relationship between GPA and the five subscales of critical thinking was significantly positive (Lombardi & Kowitt, 2014). Similarly, for these students, there were significant correlations between the majority of the critical thinking subscales and PSAT/NMSQT Math, Writing, and Critical Reading. However, the relationship between academic factors and critical thinking looked very different for students with disabilities. For students with disabilities, the relationship between GPA and critical thinking was significant for only two subscales (Problem Formulation and Research) (Lombardi & Kowitt, 2014). Most importantly, the diagnostic perspective on test-taking should be able to accurately reflect areas of strength and areas of improvement in both students without and students with special needs. Standardized tests, like final grades, provide no information that can be used as feedback by the student regarding how to improve, and some have posited that the tests do not have enough questions to adequately assess any one skill (Tienken & Orlich, 2013). Unfortunately, according to the study performed by Lombardi, Kowitt, and Staples (2015), students with special needs are not performing well on standardized tests.

In fact, Dintersmith and Wagner (2015) provide, in extreme detail, a more appropriate test that assesses cognitive rigor. This test is Arom and Roksa's Collegiate Learning Assessment (CLA), administered by the Council for Aid to Education. The CLA uses complex, open-ended problems and relevant background materials, compared to tests such as the SAT and PSAT's use of multiple-choice questions. Dintersmith and Wagner (2015) claim that the CLA measures the skills that higher education proposes to teach, including critical thinking and problem-solving. To that end, Zhao introduces the idea that Program for International Student Assessment (PISA) test-taking nations are at risk of losing creativity and the entrepreneurial spirit due to excessive focus on PISA test scores. Zhao proposes that China, for example, could never produce an innovation legend like Steve Jobs because of the focus on rote memorization, the narrow tracks students are placed in, and the sociological constraints placed on a student in China. Each of the previously listed factors plays a large part in preventing the creativity, critical thinking, transfer of knowledge, and meaning-making required for students and, eventually, employees and employers from solving open-ended, complex problems that exist in the real world. In any professional setting, decisions are made on a daily basis without the correct answer. Working in a doctor's office or a mechanic's garage is not an exercise in rote memorization-the experience is a diagnosis of a problem with a content based on knowledge applied to a new and unprecedented problem on a daily, if not hourly basis (Zhao, 2012). The success on the PISA exam has an inverse relationship with entrepreneurship activities. Zhao claims that countries that rank high on the PISA, such as Korea and Japan, are not generally well-known for producing creative innovators:

> According to the 2010 Global Entrepreneurship Monitor (GEM) report, out of the 22 innovation-driven (developed countries), which China is not part of and Singapore did not participate in the Ginkgo Evaluation of Memory (GEM) study, Korea and Japan were at the bottom, taking 19th and 21st, respectively, in terms of nascent entrepreneurship rate, or percentage of people actively seeking to establish businesses in the next three years. In other words, countries that have higher PISA scores have lower entrepreneurship activities. Specifically, those countries that show better performance on the PISA tend to have fewer people who intended or plan to start businesses and fewer people who have started businesses. (Zhao, 2012)

According to Neal, who was interviewed by Dintersmith and Wagner (2015), the innovation-economy business models eschew the high-achieving students in today's schools because of the lack of communication, collaboration, creativity, and critical thinking. Invariably, the underpinning of high-stakes testing is a person with a modus operandi that can download information, regurgitate information, and then purge information. Consequently, interpersonal skills, habits of mind, and the ability to solve an open-ended, complex problem seem to be an area that needs to be improved.

Disappointingly, through the lens of Dintersmith and Wagner (2015), the education system has crumbled under the pressure of high-stakes testing, suggesting that these tests hamper creativity and hurt students' motivation.

> Today, assessment in our schools has become the bitter enemy of learning. It is perverting the school agenda. It is killing curiosity and motivation. It is driving our best teachers from the profession. Assessment in our schools has become the single biggest threat to our nation's long-term national security. It is corroding our nation's education society in the same way invasive species like kudzu or snakehead fish drive healthy species from our environment. (Dintersmith and Wagner, 2015)

The authors go on to give a voracious description of the SAT's purpose:

> The SAT is not designed to determine whether a student has mastered important or essential skills. It is designed to generate a bell curve of results… What's so interesting

> about the bell curve's role in standardized tests is that it doesn't reflect anything real. *It was chosen arbitrarily by the College Board*. As the College Board itself explains in a report, 'So we chose the normal distribution for its symmetry and familiarity. We did not choose it because we believe ability is normally distributed.' Simply put, the statisticians creating the Scholastic Aptitude Test designed the outcomes to represent a bell curve because the bell curve is familiar. The creators of the test place arcane and confusing prompts within the test to push the tails of the bell curve outward, and allow the flurry of test scores to settle into a neatly curated bell curve. The other strategy used to create the "normal distribution" of test scores is the density of the questions per each section. In essence, the quantity of questions in a short amount of time limits the number of questions students can answer correctly. Overall, the process of creating questions is studied, re-studied, and churned over to provide the SAT with the Consequently, the difference between a 700 and a 600 is a few confusing questions that separate students from elite post-secondary institutions. Sadly, the lack of creativity of the designs of the tests has instituted a frantic sprint for students to overindulge in test preparation or work themselves into a frenzy from the anxiety of the test results. (Dintersmith & Wagner, 2015)

To prepare for these tests, students review questions similar to:

Figure 1. Sample question from Most Likely to Succeed, Dintersmith and Wagner (2015).

There is no doubt that Larry is a genuine ______: he excels at telling stories that fascinate his listeners.

a. Braggart
b. Dilettante
c. Pilferer
d. Prevaricator
e. Raconteur

Students are challenged to read passages that may not be clear and seem to angle paragraphs to be purposefully confounding as students attempt to answer questions that can look like the example below:

Figure 2. Sample question from Most Likely to Succeed, Dintersmith and Wagner (2015).

The author's tone in this passage is best characterized by the adjective:

a. Bold
b. Endogenous
c. Perfidious

d. Voracious
e. Audacious

Dintersmith and Wagner (2015) also talk about the lack of contextualization and operationalisation of mathematical concepts with the high stakes standardized testing seemingly looking for students to simply apply knowledge to equations in a math section rather than synthesize or analyze multiple sets of data to craft or justify a data-based argument:

Figure 3. Sample question from Most Likely to Succeed, Dintersmith and Wagner (2015).

If $x > 1$ and $\sqrt{x} / x = x^m$, what is the value of *m*?
a. -7/2
b. -3
c. -5/2
d. -2
e. -3/2

Dintersmith and Wagner (2015) express their concern with the amount of value placed on an SAT score when the SAT is only loosely correlated with anything of consequence. Most importantly, high-stakes standardized testing does not equate to mastering an important life skill (Dintersmith and Wagner, 2015). The question this research study is looking to answer is what skills does the PSAT/NMSQT or any standardized test really examine?

Lastly, a qualitative investigation into the perceptions of students graduating from a prestigious and rigorous college illuminates the disconnect between the standardized test scores and higher-order thinking skills that plays a critical role in college and career readiness. The Massachusetts Institute of Technology (MIT) is widely regarded as the world's preeminent engineering educational institution, with a long history of achievement and truly outstanding alumni. Of past Nobel Prize winners, some eighty-one have ties to MIT. In a typical year, researchers at MIT file hundreds of patents, start dozens of technology companies and publish hundreds of scholarly articles. Most people view MIT as being the pinnacle of engineering education. A few years ago, an MIT student, Kristen Wolfe, focused her thesis on the topic of "Understanding the Careers of the Alumni of the MIT Mechanical Engineering Department." She looked at a wide range of things students could learn and use, whether at MIT or otherwise. She surveyed over three hundred graduates of MIT's Mechanical Engineering Department who were, at the time, about a decade into careers spanning academics, research, business, and nonprofits.

Respondents to Wolfe's survey reported that they spent almost all of their time at MIT on narrow topics that they would not end up using professionally (e.g., fluid mechanics, heat transfer, obscure mathematics) and almost no time at MIT on essential skills professionally. Wolfe, as reported by Dintersmith and Wagner (2012) concludes, "The largest disconnects are in the areas of personal skills, professional skills, independent thinking, teamwork, and communication. These areas received the largest scores for proficiency and frequency and the lowest for learning at MIT." MIT is among the most prolific academic institutions in the world, and the students who studied there stated that they were at a deficit in their careers because of the lack of interpersonal skills and independent thinking. Just to frame the prestige of MIT: the incoming class of 2023 included 21,312 students who applied, and 1,427 were admitted (6.7%). According to MIT admissions, the typical student who is accepted to the world-renowned Massachusetts Institute of Technology is an incredible example of standardized-test taking prowess. Invariably, the admissions statistics of students admitted are nothing short of outstanding. Each of the following tables containing MIT admissions data have been adapted from https://mitadmissions.org/apply/process/stats.

Figures 4-9. Admissions data tables adapted from Massachusetts Institute of Technology (2020).

TEST	**RANGE**
SAT Math	[790, 800]
SAT ERW	[730, 780]
ACT Math	[35, 36]
ACT English	[35, 36]
ACT Composite	[34, 36]

Distribution of SAT scores (Math)

SCORE	**APPLICANTS**	**ADMITS**	**ADMIT RATE**
750-800	12,679	1,065	8%
700-740	1,726	24	1%
650-690	814	1	0%
600-640	399	0	0%

<600	318	0	0%

Distribution of SAT scores (ERW)

SCORE	APPLICANTS	ADMITS	ADMIT RATE
750-800	6,504	677	10%
700-740	5,039	312	6%
650-690	2,614	87	3%
600-640	1,091	11	1%
<600	688	3	0%

Distribution of ACT scores (Composite)

SCORE	APPLICANTS	ADMITS	ADMIT RATE
34-36	5,848	569	10%
31-33	1,689	73	4%
28-30	590	7	1%
25-27	296	1	0%

<25	205	0	0%

Distribution of ACT scores (Math)

SCORE	APPLICANTS	ADMITS	ADMIT RATE
34-36	5,882	583	10%
31-33	1,524	67	4%
28-30	697	0	0%
25-27	352	0	0%
<25	173	0	0%

Distribution of ACT scores (English)

SCORE	APPLICANTS	ADMITS	ADMIT RATE
34-36	6,425	590	9%
31-33	1,015	39	4%
28-30	512	14	3%
25-27	341	6	2%

<25	335	1	0%

The students that apply to MIT (including the students who do not get admitted) have some of the most prolific test scores in the nation and across the globe and, incredibly, the alumni of this excellent institution still believe that the skills they developed at MIT may have been insufficient for their future careers.

Benefits of High Stakes Standardized Testing

On the other hand, high-stakes, standardized tests such as the PSAT/NMSQT and the SAT have been found to have strong benefits for the test-takers as well. The Standardized Test Task Force of the University of California system [STTF] and its Writing Subcommittee consulted dozens of studies concerning standardized tests, their predictive value, and their impact on access and diversity. The STTF met with the national testing agencies, critics of standardized testing, State education leaders, University of California (UC) campus admissions officers, University of California Office of the President (UCOP) institutional researchers, Board of Admissions and Relations with Schools (BOARS), and other UC-based and non-UC content experts.

> The STTF found that standardized test scores aid in predicting important aspects of student success, including undergraduate grade point average (UGPA), retention, and completion. At UC, test scores are currently better predictors of first-year GPA than high school grade point average (HSGPA), and about as good at predicting first-year retention, UGPA, and graduation. For students within any given (HSGPA) band, higher standardized test scores correlate with a higher freshman UGPA, a higher graduation UGPA, and higher likelihood of graduating within either four years (for transfers) or seven years (for freshmen). Further, the amount of variance in student outcomes explained by test scores has increased since 2007, while variance explained by high school grades has decreased, although altogether does not exceed 26%. Test scores are predictive for all

> demographic groups and disciplines, even after controlling for HSGPA. In fact, test scores are better predictors of success for students who are Underrepresented Minority students (URMs), who are first-generation, or whose families are low-income: that is, test scores explain more of the variance in UGPA and completion rates for students in these groups. One consequence of dropping test scores would be increased reliance on HSGPA in admissions. The STTF found that California high schools vary greatly in grading standards, and that grade inflation is part of why the predictive power of HSGPA has decreased since the last UC study. (University of California Academic Senate, 2020)

The University of California's STTF put in an extraordinary amount of time, energy, and effort to provide a measured, informed, and statistical analysis of the high-stakes, standardized testing, and to determine how the results can predict a student's success at the collegiate level. The report created by the STTF delivers a wealth of information. One of the more significant points the STTF report raises is how test scores might affect the student experience for underrepresented minority students. In a day and age where social justice is deservedly at the forefront of education, this fact should not go unnoticed! Moreover, the valid predictive nature in which an SAT or an ACT accounts for first-generation students and students who have a background of a lower socioeconomic status is also of the utmost importance. Additionally, the high-stakes standardized testing seems to level the playing field in access to higher education, but the data presented by the STTF on standardized tests indicate a higher predictor of first-year retention rate for students. The students are not only gaining access to higher education but also remaining on campus after their first year, which is a clear sign of successful commitment to academics. Furthermore, the STTF research found that, even among students with similar high school credentials, there is a significant relationship between test scores and college retention, graduation rates, and grades (University of California Academic Senate, 2020). Furthermore, for any given high school GPA, a student admitted to a college or university that scored low on the

SAT has been found to be between two and five times more likely to drop out after their freshman year, and up to three times less likely to complete their degree compared to a student that scored highly on the SAT (University of California Academic Senate, 2020). Consequently, one could claim the SAT and the PSAT/NMSQT support access to higher-level institutions. The data reviewed in the STTF's research supports the idea that the SAT is predictive of students' performance once they arrive on campus.

The research performed by STTF also found that the SAT does indeed have the ability to propel underrepresented high school students to receive admission in the University of California collegiate system:

> The Task Force found that of the 22,613 students guaranteed admission through the statewide index (the admissions pathway in which test scores can compensate for lower HSGPA) but not through Eligibility in the Local Context (ELC), which only considers high school grades, about 25% were members of underrepresented minority groups, and 47% were low-income or first generation students. (University of California Academic Senate, 2020)

Inherently, the SAT allowed students that may otherwise not have had a chance to receive a higher-level education to gain admission to a college education. Significantly, the SAT was originally created to identify students that did not have high high-school GPAs or did not attend a well-funded high school but who still demonstrated the skills needed to succeed in college:

> Many students who do not qualify for ELC, (i.e., being in the top 9% of their classes based on HSGPA), become eligible to attend UC through statewide eligibility. Unlike ELC, statewide eligibility, which is also referred to as Index Eligibility, uses the SAT as well as GPA. (University of California Academic Senate, 2020)

The University of California Institutional Research and Academic Planning (IRAP) also studied the predictive power of grades and test scores on students' college grades in different majors. Notably, the STTF discovered through their study that the SAT was best at predicting academic performance in the quantitative disciplines. Additionally, this analysis disaggregated freshman GPA at the course level, which allowed for a more precise examination of the effects of test scores and HSGPA in explaining student performance in different academic disciplines. They found that the GPA-only models in their regression analyses were poor predictors of college performance but, when combined into a model that included SAT scores, HSGPA and SAT scores combined were able to better predict college grades. They considered this an indication that SAT scores do, in fact, provide information about prospective student performance (University of California, Academic Senate, 2020).

Resultantly, the research, time, and energy put forth by the University of California Systemwide Review of the Report of the Academic Council's Standardized Testing Task Force have shown that the SAT and the PSAT/NMSQT's claim of being a predictor for career and college readiness could be substantiated through their data. An article published in Capitol Weekly affirms the findings from the STTF:

> First, since the College Board and our members redesigned the SAT in 2016, it now tests what students actually learn in high school. Gone are the "SAT words," penalties for getting questions wrong, and tricky math questions. Secondly, there used to be a huge inequality in students' ability to practice for the SAT… But in 2015 the College Board tackled the commercial test prep industry head-on by partnering with another not-for-profit, Khan Academy, to create a completely free online, personalized practice tool (satpractice.org) that nearly 10 million students have used to practice for the SAT. Thirdly, students used to face real barriers when taking the SAT on a Saturday, especially those who worked to help their family or who lacked transportation to a weekend testing site. Today, over 40% of all SAT administrations — taken by nearly one million students in the class of 2019 — are held at the students' school during the school day, by all students and at no cost to them. When more students are able to take college admissions tests like the SAT, more underrepresented students can be connected directly to scholarships, free test practice tools, and college application fee waivers. And lastly, we heard from many lower-income and first-generation students that they felt completely overwhelmed by the college planning and application process… in 2018 we created a $25 million scholarship program—the College Board Opportunity Scholarships—that gives students a college application plan and rewards them

> with scholarships just for taking important college planning steps, including practicing for the SAT on Khan Academy, building a college list, completing the FAFSA for financial aid, and applying to college. (Rodriguez, 2020)

In essence, the peripheral perks of participating in standardized tests like the PSAT/NMSQT, the SAT, or the ACT open doors to students who may be economically disadvantaged that otherwise may be locked out of receiving higher education.

The PSAT/NMSQT's parent company, the College Board, has worked to provide data and research throughout the test's lifetime to ensure the validity and reliability of the PSAT/NMSQT. In fact, the College Board performed an internal review to construct a solidified list of PSAT/NMSQT Indicators of College Readiness. The College Board's team of researchers found that the College Readiness Index introduced by Wiley et al. (2010) had predictive ability to determine whether students were well prepared for college-level work; however, they note that the information needed to complete the index is not available for most students do not provide until spring of their junior year or later. Significantly, one of the benefits of the PSAT/NMSQT is the opportunity for students to see if they are on track for college, and by correlating PSAT/NMSQT scores to the subsequent SAT scores, sophomores, juniors, and then seniors can reflect on their college and career readiness (Wiley et al., 2010). Subsequently, according to the study conducted by Proctor, Wyatt, and Riley (2010), the PSAT/NMSQT provides valuable information as a formative assessment to help students, teachers, administrators, parents, and stakeholders understand the status of a students ability to be proficient in their first year of college at a young age. If students were to take the PSAT/NMSQT as a sophomore or junior in high school, they would be able to remediate their areas of weakness and accelerate their areas of strength in a timely fashion to continue to propel themselves on a path to college readiness. Furthermore:

> The PSAT//NMSQT also must respond to and refute commonly held misconceptions about tests and their uses, including whether scores at the top end of the scale meaningfully discriminate among test-takers. In an article reviewing commonly held test misperceptions, Sackett, Borneman, and Connelly (2008) pointed out that there is a misconception that tests are unable to distinguish among high scorers and are only useful for weeding out low scorers. However, research has shown that the ability–performance relationship is linear throughout the entire range of ability, across various domains. This linear relationship provides evidence that tests can distinguish among different ability levels equally and are not only useful for weeding out low-scoring individuals. In other words, with each increase in test scores, there is an associated increase in performance even among very top performers. (Marini, Mattern, & Shaw, 2011)

Marini, Mattern, and Shaw (2011) were contracted by The College Board to statistically analyze the relationships between PSAT/NMSQT Selection Index (which includes all components of the PSAT/NMSQT: Reading, Math, and Writing) scores and first-year grade point averages of college students. The study was additionally focused on the relationship between high-scoring students and the predictability of their first-year grade point average. The study concluded that “higher PSAT/NMSQT scores are related to higher grades earned in college, supporting its use as a screening tool in the overall process of selecting Merit Scholarships” (Marini, Mattern, & Shaw, 2011). Again, as a predictor of first-year college success, the PSAT/NMSQT seems to be a valid indicator of future success for students. In addition, prior research has demonstrated the utility of the SAT and High School Grade Point Average (HSGPA) in predicting Freshman Year GPA (FYGPA). Kobrin, Patterson, Shaw, Mattern, and Barbuti (2008) investigated the predictive validity of the SAT and found that the SAT sections combined correlated .35 with FYGPA, but when correcting for restriction of range, the correlation was .53 (adjusted $r = .53$). HSGPA had a similar correlation of .36 with FYGPA (adjusted $r = .54$).

Together, SAT and HSGPA correlated .46 with FYGPA (adjusted $r = .62$), and thus the results suggested that the combination of SAT and HSGPA was the best predictor of Freshman Year GPA.

The ability to easily track the students' college readiness has been a goal for many researchers interested in school reform (Center for American Progress, 2009). School districts across the United States are supplied with the post-test metrics. At times, the data from standardized testing are organized, correlated, and organized to help schools study subgroups of students (e.g., according to ethnicity, socioeconomic status) to help identify patterns within the data and better serve underprivileged populations. One of the identified issues raised to the College Board in reference to their presentation and organization of the data is the lack of reference points to analyze the data longitudinally. Subsequently, the Multidimensional College Readiness Index "delineates likely college ready from not likely college ready, to help educators and policymakers determine how many of their students are actually prepared to succeed in college" (Wiley, Wyatt, & Camara, 2010). The study conducted by Wiley, Wyatt, and Camara (2010) had a primary objective to create a metric that could incorporate multiple data points with an SAT score outcome. Other factors involved in the Multidimensional Index include high school GPA and the difficulty of the classes the student has completed during their high school career. Ultimately, the goal is to involve the College Board in assisting local school districts to increase the rate of college readiness among their learning community through the lens of the AP classes, the PSAT/NMSQT, and the SAT.

Significantly, in terms of the consequences of learning, practicing, and sharpening higher-order thinking skills before students graduate high school, Wiley, Wyatt, and Camara

(2010) detailed how having a bachelor's degree can promote a person in the United States to higher earnings over their career, higher employment rates, more activity in volunteerism, higher job satisfaction, lawful behavior, and life satisfaction. Obtaining a Bachelor's degree has largely become a baseline requisite for employment in the United States of America. Invariably, colleges and universities look to capitalize on the expediency in which students graduate as well. Unfortunately, a portion of students entering colleges and universities are compelled to enroll in remedial courses to help propel those students to successfully complete their four-year degrees (Kelly & Ewell, 2009). Expectedly, various colleges have invested time and resources to help identify the skills and knowledge that enable students to be more college ready. The reported outcomes of Wiley, Wyatt, and Camara's study (2010) concluded that one out of three students included in their study were college ready. The researchers anticipated about 50% of students to reach at least one of the "college readiness" benchmark scores on high-stakes standardized testing. The results suggested that one out of five students were not reaching any of the three benchmarks. This extensive study not only provided data regarding the SAT's ability to help students become college readiness but also shed light on the unpreparedness of students achieving that college readiness. While the proposed index can provide useful information for schools and districts to evaluate the students graduating from their high schools, it can also be used as a benchmark to develop a series of early indicators of college readiness. Many researchers who have focused on college readiness have encouraged the development of clear college readiness indicators for students as they progress through school (Corwin & Tierney, 2007). Additionally, through the development of the ARI (Academic Readiness Indicator), a clear set of course recommendations could be developed that would allow schools and districts to track what percentage of their students were taking an appropriate level of course work each year. The ARI could allow educators to compare the rigor of courses across groups of students and across years to determine if and where significant gaps exist and how to respond to these needs. Finally, schools can examine gaps and weaknesses (e.g., test scores, grades or courses, as well as specific course areas) and evaluate the impact that potential changes in course sequence, access to honors, Advanced Placement courses, and curriculum may have on the number and percentage of students who are college ready in high school. With substantial research, similar projections could also inform projections of college readiness in the eighth and ninth grades (Wyatt, Wiley, & Camara, 2011). The utility of the educational apparatus that the College Board has created in a concerted effort to prepare students for college readiness is thorough, well-funded, and studied by some of the field's most proficient researchers. The entire College Board apparatus has a structure that guides a prospective college student starting in the eighth grade through the formative years of high school. Accordingly, the College Board has access to an arsenal of resources from all of the College Board products and is able to illustrate the data of a student's college readiness in partnership with local school districts to help prepare that student for postgraduate studies.

Proctor, Wyatt, and Wiley (2010) posit that their proposed index as correlated to student performance on the PSAT/NMSQT in the 10th and 11th grades, students who need remedial work on college and career readiness can better be identified at earlier stages in their academic careers based upon their PSAT/NMSQT scores. Potential applications and benefits of these indicators associated with the PSAT/NMSQT and the SAT may also be useful to schools, districts, and states that may need assistance in developing and validating their own measures of college readiness.

> In the first analysis, benchmark scores for 10th- and 11th-grade PSAT/NMSQT test-takers were created. In the case of the 11th-grade PSAT/NMSQT benchmark scores, logistic regression was used to obtain the minimum junior PSAT/NMSQT score associated with a 65 percent probability of obtaining the SAT college readiness benchmark In the second analysis, contingency tables were established to show the percentage of students who went on to meet or exceed the SAT college readiness benchmark by PSAT/NMSQT score band. The percentage of 11th-grade PSAT/NMSQT students who met or exceeded the SAT benchmark was calculated for each 5-point PSAT/NMSQT score band The last step of the analysis involved using data from the 2009 PSAT/NMSQT test to evaluate the impact of the derived benchmarks on the population of PSAT/NMSQT test-takers. Benchmarks obtained in the first analysis were used in this step. This analysis was performed to validate the benchmarks determined from the previous analysis. (Proctor, Wyatt, & Wiley, 2010)

The data-driven indicators based on PSAT score results created by Proctor, Wyatt, and Wiley could present a useful summary of students' college readiness as they leave high school and could be a valuable component of a district's overall college readiness toolbox.

The College Board Standards for College Success define the skills and knowledge students must develop and master to succeed in college and the workforce in the 21st century. These standards for English Language Arts, Mathematics and Science are based on empirical research conducted by the University of Oregon's Center for Educational Policy Research in collaboration with the Association of American Universities. The standards are benchmarked against the College Board's Advanced Placement Program® as well as national and international frameworks including National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA). The College Board Standards are designed to:

> Provide a model set of comprehensive standards for rigorous middle school and high school courses that lead to college and workplace readiness, reflect 21st-century skills such as problem solving, critical and creative thinking, collaboration, and media and

> technological literacy, serve as both learning and performance standards to guide curriculum development, instruction, and formative and summative assessment development, and provide teachers, districts and states with tools for increasing the rigor and alignment of course work across grades 6–12 to college and workplace readiness. Several states have used the College Board Standards as a resource or benchmark in reviewing their own state standards for alignment to college readiness and identify strengths and areas needing improvement. (College Board, 2020)

Overall, the definition of college readiness used in this report is tied to the academic success of students. Students who are college ready should be able to succeed in entry-level, credit-bearing college courses without the need for remediation. Other factors associated with college success (e.g., motivation, study skills, attitudes) may be equally important in evaluating outcomes such as persistence and college graduation or post-college career readiness. Yet this more expansive definition of college readiness is not the focus of this report (Wyatt, Wiley, & Camara, 2011). Essentially, the standardized test score results in related research to the substance of the PSAT/NMSQT has been both positive and negative for students, administrators, stakeholders in the field of education, families of different socioeconomic backgrounds, students with different abilities, and teachers. But how exactly should the PSAT/NMSQT, the preliminary career, and college readiness test, be testing high school students?

College and Career Readiness

The concept of College and Career Readiness (CCR) began to take shape in the early 2000s. It became formalized for policymakers, educators, and administrative team members when Conley (2010) published *College and Career Ready: Helping All Students Succeed Beyond High School*. In fact, as part of the Bill and Melina Gates educational initiative to provide access to high-quality education for all students regardless of socioeconomic background they partnered with Dr. Conley. Correspondingly, "several studies have found college faculty members nationwide regardless of the selectivity of the institution, to be in near-universal agreement that most students arrive unprepared for the intellectual demands and expectations of postsecondary faculty" (Conley, 2010). Consequently, Conley investigated and summarized the elemental skills for students to be placed in a position to succeed after they graduate from high schools in the United States. Based upon the research of Conley (2010), college and career readiness is the level of preparation a student needs to enroll in and succeed in college or university, without needing remedial courses, with *succeed* in the definition referring to the student mastering the skills and knowledge needed to then take the next course in the sequence on the way to receiving a diploma or completing a certificate.

Centrally, the college and career ready student is able to grapple with open-ended, complex problems and establish new cognitive pathways as new information and skills are presented to their academic or professional repertoire. Accordingly, Conley also details the four dimensions of college and career readiness:

> Key cognitive strategies, key content knowledge, academic behaviors, and contextual and awareness skills. Other factors exist outside this model, such as student motivation and family support, but the four dimensions of this model encompass all of the areas for which high schools can reasonably be expected to take primary responsibility to provide all students the necessary learning experiences and programs of preparations. (Conley, 2010)

Placing college students in a position to succeed is contingent upon the student's ability to understand, recall, and apply new information across the curriculum:

> The term key cognitive strategies was selected deliberately to describe the intentional behaviors students must be able to employ situationally and to emphasize that these behaviors need to be repeated frequently and over an extended number and variety of situations so that students learn when and where to employ them. (Conley, 2010)

The proposed key cognitive strategies are listed below:

- Problem formulation
- Research
- Interpretation
- Communication
- Precision and accuracy

In summation, the cognitive strategies created by Conley (2010) were carefully chosen as skills that can and will help students succeed in whatever postsecondary path they choose. Further, Conley continues his structure of career and college readiness as he details the traits of Key Content Knowledge, which is the application of knowledge to propel the student's understanding on a deeper level with an eventual goal of expertise and the ability to transfer this knowledge among disciplines in a constructive manner (Conley, 2010). Additionally, the student's individual behavior has an influence on their development in their occupational or academic future: Self-monitoring, or the ability to think about one's own cognitive performance, understanding how one learns, processes, and applies new material, is described as an additional, necessary skill for academic success (Conley, 2010). Conley (2010) also mentions that students must be able to persevere and stay on track while maintaining a commitment to their mental health, the positive relationships among those around them, and other environmental factors that may arise as challenges to the growth of a student.

Notably, skills required to succeed beyond high school and college are almost similar to the College Board's definition of college readiness in terms of what the PSAT/NMSQT allegedly assesses. Specifically, the research questions of this very study intend to identify career and college readiness skills by asking: does the language found on the grade 10-11 PSAT/NMSQT

Reading Section questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking. Conley's (2010) approach to Career and College readiness links to the concept of higher-order thinking skills.

Conley has worked extensively within both academia and public school districts to ensure that children in the United States of America are receiving a proper education, with higher-order thinking skills being taught, practiced, and assessed with a high enough standard of cognitive rigor that children can be successful for whichever path they choose following high school graduation. As the researchers have detailed throughout the literature review, the concept of higher-order thinking skills, 21st-century learning, and cognitive rigor is no secret. Yet, it seems to elude most educators and, unfortunately, most students as well. Other researchers have examined how little is being learned in college. Keeling and Hersh (2011) illustrate how a higher quantity of college graduates fail to sharpen critical thinking skills, creativity, communication, and the ability to adapt or even be a contributing member of a team. A bachelor's degree does not indicate a set of functional workplace skills or interpersonal skills that can be utilized in the modern office space (Keeling & Hersh, 2011). It has been estimated by the National Center for Education Statistics that 20 percent of college graduates struggle with even the most basic quantitative skills, and an additional one-third of college graduates struggle with reading and comprehending complex books (Dintersmith & Wagner, 2015).

> College is expensive, more so each year. Applications continue to soar, particularly to elite institutions. So it would seem logical to conclude that colleges do an amazing job of educating students. But do they? ...the majority of college students learn little to nothing on the important dimensions of critical thinking and analysis, complex reasoning, and writing. (Dintersmith & Wagner, 2015)

In an exhaustive approach to study and measure higher-order thinking skills, *In Search of Deeper Learning* (2011) by Mehta and Fine performed an incredible analysis of the higher-order thinking skills practiced by students at high schools across the United States compared to international secondary schools. Mehta and Fine's goal was to study the varieties of approaches to deeper learning that schools were taking and to learn from those that were doing the best at this task. The overview of the six-year study found that the United States was lagging in providing teachers and students with chances to practice cognitive rigor. Fundamentally, the study resulted in about "four out of five classrooms Mehta and Fine visited featured tasks that were in the bottom half of Bloom's taxonomy, asking students to recall, comprehend, or apply, rather than to analyze, synthesize, or create" (Mehta & Fine, 2011). This was:

> The largest ever videotaped study of American classrooms, grades four through eight, in which more than seven thousand videos were scored by multiple observers across four different validated instruments, found that American teachers scored high in 'behavior management' but were weakest at 'analysis and problem solving, regard for student perspectives, quality of feedback and content understanding.' Of these competencies, 'analysis and problem solving' was the least frequently observed, seen in only about 20 percent of lessons (similar to our one in five estimate). In math, only 1 percent of lessons scored in the top rating for analytic complexity, while 70 percent received the lowest rating. Similar findings were reported in a large-scale analysis by Education Trust, which focused on the tasks that middle-school students are asked to carry out. In evaluating nearly 1,600 tasks at six middle schools, analysts used Webb's depth-of-knowledge scale to examine the complexity of the tasks that students were asked to complete. They found that only 4 percent of assignments asked students to think at higher levels. (Mehta & Fine, 2011)

Moreover, Mehta and Fine (2011) discovered through their observations that the teachers frequently lost out on opportunities to push students to develop higher-order thinking skills and develop the students' cognitive rigor. Unfortunately, most teachers observed limited student responses to frame their own thoughts during the study, and the Socratic seminar approach was found to be of limited interaction between student peers. Mehta and Fine (2011) also discovered that the most captivating educators could nimbly and artistically dance between lower levels of Bloom's taxonomy and higher level of Bloom's taxonomy as they continuously pushed students to develop, practice, and reflect upon their own higher-order thinking skills. To illustrate this concept even further, Mehta and Fine (2011) dove into the realm of high-stakes standardized testing, noting that teachers, as a result of the focus on standardized tests, feel obligated to move quickly to cover the vast curriculum, but not nearly deep enough to allow for students to sharpen their higher-order thinking skills in a meaningful fashion. The pressure to cover the curriculum constructed a vicious cycle in which the students who want to do good at school comply with the teachers low-level assignments, and the teachers who want to cover the breadth of material without any depth move forward without any pause for the students to reflect, grapple, debate, make connections or synthesize the learning experiences.

In the final analysis portion of the book, Mehta and Fine detail the ideal future of schooling through the lens of the pursuit of deeper learning, and thus propelling our students to new heights after high school graduation. One specific section strategizes about a necessary shift in assessment:

> Rather than evaluating students on seat time in conventional classes or on the results of standardized tests, the goal would be to evaluate students on the basis of their performance in domain-specific areas. [For example,] The 'assessment of the students in *The Servant of Two Masters* was their show night; the assessment of students who produce the student newspaper is the quality of that newspaper; the assessment of project-based learning is the quality of the final artifact that is presented or performed. (Mehta & Fine, 2011)

Hence, the students are required to use higher-order thinking skills to display their learning in an authentic utilization of the tools (e.g., skills, knowledge, innovation, and synthesis) they have acquired by a high-quality education. An increasing number of higher education initiatives focusing on measuring and improving college readiness of high school students have been constructed due to the increase of remedial measures at the college level to keep students on track for graduation (Conley, 2010). One of the more notable efforts to define college readiness was the project, Understanding University Success, sponsored by the Association of American Universities and the Pew Charitable Trusts (Conley, 2003). This two-year project recruited more than 400 faculty members from 20 research universities. Faculty members identified the knowledge and skills that students needed to have in order to succeed in entry-level courses at their universities. Faculty members were recruited across all disciplines, and standards were developed in English, mathematics, natural sciences, the social sciences, second languages, and the arts. Beyond knowledge and skills required in each discipline, the authors noted that habits of mind are critical for college success (Conley, 2003). The habits of mind discussed included critical and analytical thinking, problem-solving, inquisitiveness, and the initiative to take advantage of the resources at their attending university. Other key features included openness to trying new things and being willing to fail at tasks the first time and the ability to accept critical feedback and adjust accordingly. These habits of mind are a critical part of the standards, and the authors note that the content knowledge included in the standards cannot be interpreted without also considering these critical cognitive skills.

The Educational Policy Improvement Center (EPIC), in conjunction with the Bill and Melinda Gates Foundation (Conley, 2007), released a second influential report, which expands the definition of college readiness and views habits of mind as being one of four concentric levels required for success. The other three areas are academic knowledge and skills, academic behaviors, and contextual skills and awareness. The habits of mind and academic knowledge and skills discussed here are typically similar to those used in Understanding University Success (Conley, 2003); however, this report also discusses the academic behaviors and the contextual skills necessary for college readiness. The academic behaviors are primarily focused on self-awareness and monitoring, as well as on study skills. Self-monitoring is essential because it allows students to work their way through a topic area independently and to determine if they have reached a comprehensive understanding of a topic. Study skills represent a wide range of key behaviors, such as time management, note-taking, and other essential skills students must have in order to navigate their way through college successfully. The contextual skills and awareness focus on students' ability to understand the university system as a whole and their role within the university. Without this, students may never understand how the university system works and find themselves frustrated at their inability to navigate the requirements and obstacles in their way (Wyatt, Wiley, & Camara, 2011).

In a study conducted by Lombardi, Kowitt, and Staples (2015), an examination of critical thinking skills associated with College and Career Readiness of high school students with and without disabilities took place. Additionally, the researchers also included several other demographic factors, including race, socioeconomic status, and English language learner (ELL) status (Lombardi, Kowitt, & Staples, 2015). "Critical thinking skills included problem formulation, research, interpretation, communication, and precision, and accuracy (Lombardi et al., 2013). In college courses, students are expected to apply this sequence of critical thinking across content areas regardless of academic discipline. For example, in a freshman writing course, college students have to formulate a thesis or "problem" statement, research the problem, find and interpret evidence, and then convincingly communicate the findings and conclusions back to an audience in different formats (e.g., essay, presentation) with minimal errors (Conley, 2007). Similarly, in a freshman science course, students have to formulate a hypothesis, test it by collecting data, interpret these data, and communicate findings and conclusions to an audience with few to no errors. Thus, despite the very different academic disciplines, college freshmen are expected to think critically as soon as they arrive on campus (Conley, 2003), and first-time job employees are expected to quickly master job-specific knowledge, skills, and tools. For example, solving an unexpected work-related challenge or material shortage, resolving a work conflict with a colleague or a supervisor, or setting new production goals, are all skills that employers expect young adults to have (Lombardi, Kowitt, & Staples 2015). Lombardi, Kowitt, and Staples (2015) focused on thinking critically for these reasons:

> (a) the currently used and well-accepted academic indicators do not necessarily measure critical thinking; in fact, evidence suggests misalignment between academic indicators and the expectations of postsecondary instructors of what students should know and be able to do as incoming freshmen (Achieve, Inc., 2007; Brown & Conley, 2007; Brown & Niemi, 2007; Conley, 2003), and (b) minimal evidence exists on differences between secondary students with and without disabilities and their critical thinking skills as measured by Conley's five-part definition: problem formulation, research, interpretation, communication, and precision and accuracy (Conley, 2007, 2010; Lombardi et al., 2013). Critical thinking has been prioritized by the Common Core State Standards (CCSS) (NGA & CCSSO, 2010) and further validated by postsecondary instructors as an expected skill for incoming freshman to possess and use when they begin first-year courses. (Lombardi, Kowitt, & Staples, 2015)

Not only have the researchers for this study identified that students with disabilities may be at a disadvantage when being assessed for critical thinking skills, but they have also delineated evidence that disavows the connection between academic indicators and a valid instrument to test for career and college readiness skills.

Measurement Tools

The creation of the tool that will be used in this study to evaluate each question's cognitive rigor of the 2018 PSAT/NMSQT has deep roots in education. At the very foundation of Hess' Cognitive Rigor Matrix is one of the cornerstones of education in which "Bloom's Taxonomy helps teachers formulate lessons that practice and develop thinking skills over a wide range of cognitive complexity" (Bloom, 1956). Although Bloom's Taxonomy was later revised by a team of education researchers headed by Anderson and Krathwohl (2001), the overall intent of the taxonomy is dedicated to categorizing questions and activities according to their levels of abstraction. However, Bloom's Taxonomy suffers limitations when selecting test items and formulating questioning strategies because the taxonomy uses verbs to differentiate taxonomy levels — many verbs appear at multiple levels and do not clearly articulate the intended complexity implied by the taxonomy (Hess et al., 2009). A new model of rigor, depth of knowledge (DOK) assists educators in being able to measure the acumen of a student's ability to apply knowledge to a prompt or a learning exercise. The resulting combination of Bloom's Taxonomy and Webb's DOK borne cognitive rigor — and efficiently forms a comprehensive, subject-specific series of rubrics that can effectively detail rigor, thus posing a wide range of uses at all levels of curriculum development and delivery (Hess et al., 2009). Subsequently, the creation of the Cognitive Rigor Matrix allows educators and stakeholders to assess higher-order thinking skills "because cognitive rigor encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities, the Cognitive Rigor Matrix can enhance instructional and assessment practices at the classroom level as well. Superposing the two cognitive complexity six measures produces a means of analyzing the emphasis placed on each matrix intersection. As educators become more skilled at recognizing cognitive rigor and analyzing its implications for instruction and assessment, they can enhance learning opportunities for all students and across all subject areas and grade levels" (Hess et al., 2009). The innovation of Hess' Cognitive Rigor Matrix provides the researchers with the tool to accurately assess the cognitive rigor of the PSAT/NMSQT.

Significantly, the creation of Hess' Cognitive Rigor Matrix was founded upon the following principles: making connections to consolidate the learning, applying what you've learned to new situations (also known as (fa) transfer), challenging with 'just right' support (also known as 'productive struggle"), (sometimes) higher-order thinking: analyzing, evaluating, and creating, engagement, collaboration, and discourse that make thinking visible, and triggered by asking a different kind of question that shifts teacher-student roles (Hess, 2018). Significantly, Hess provides an outline of several poignant notions that cognitive rigor tends to be misconstrued by educators:

- Some educators may believe that critical thinking is an independent pursuit or even go as far to believe that some students are simply unable to think deeply. Hess believes that not only do students have an ability to grow into critical thinkers through practice and scaffolding, but Hess is a strong proponent of students collaborating as they work to sharpen their cognitive rigor skills through dialogue and exchanging ideas and strategies with peers (Hess, 2018).
- Webb's Depth of Knowledge is an organizational tool that assists educators with how students work with and interact with content. The utilization of Webb's Depth of Knowledge as a measurement tool for instructional delivery or learning experiences is not

necessarily always going to reach the highest level (Extended Thinking). As students work to acquire knowledge and skills, the learning tasks or the assessment may be Level 1 or Level 2 on Webb's Depth of Knowledge which may be appropriate for that particular class. Essentially, cognitive rigor and DOK are a balance of how you acquire and use what you've learned in increasingly more complex ways (Hess, 2018).

- Hess cautions against overly committing to verbs as the aspiration of the student outcome. Verbs describe a type of thinking, not the depth of understanding. Hess goes on to illustrate that the ultimate understanding of a skill or content is actually what comes after the verb, the content <u>and</u> the engagement with that content that helps educators determine the complexity of the task, not the verb (Hess, 2018).
- Hess is emphatic about the underpinning of cognitive rigor is the opposite of the notion that learning exercises get harder or more difficult. Quite the contrary, cognitive rigor is fundamentally focused on flexible thinking, seeing multiple possibilities, multiple approaches, and different possible perspectives. Cognitive demand is about the complexity of the task and the mental processing required to complete it (Hess, 2018).
- The belief that multiple-choice questions can access higher levels of Webb's DOK is frowned upon according to Hess' research. Invariably, the level of engagement with content that is required by DOK 3 and DOK 4 tasks are most effectively assessed by prompting higher-order thinking skills through the use of constructed-response questions, performance tasks, and extended projects. Hess concedes that Webb's DOK Level can be assessed with multiple-choice questions that ask a test-taker to determine the author's purpose, identify text evidence or data that support a claim, interpret complex graphs or

data sets, or identify flaws in an experimental design. Appreciably, selected-response and short-answer questions are an efficient means for assessing most foundational (DOK 1) and conceptual knowledge and skills (DOK 2). However, by asking a student to choose the "best option" does little to incite synthesis, connections among multiple sources, creativity, or the student's journey of metacognition that reaches Webb's DOK Level 4 (Hess, 2018).

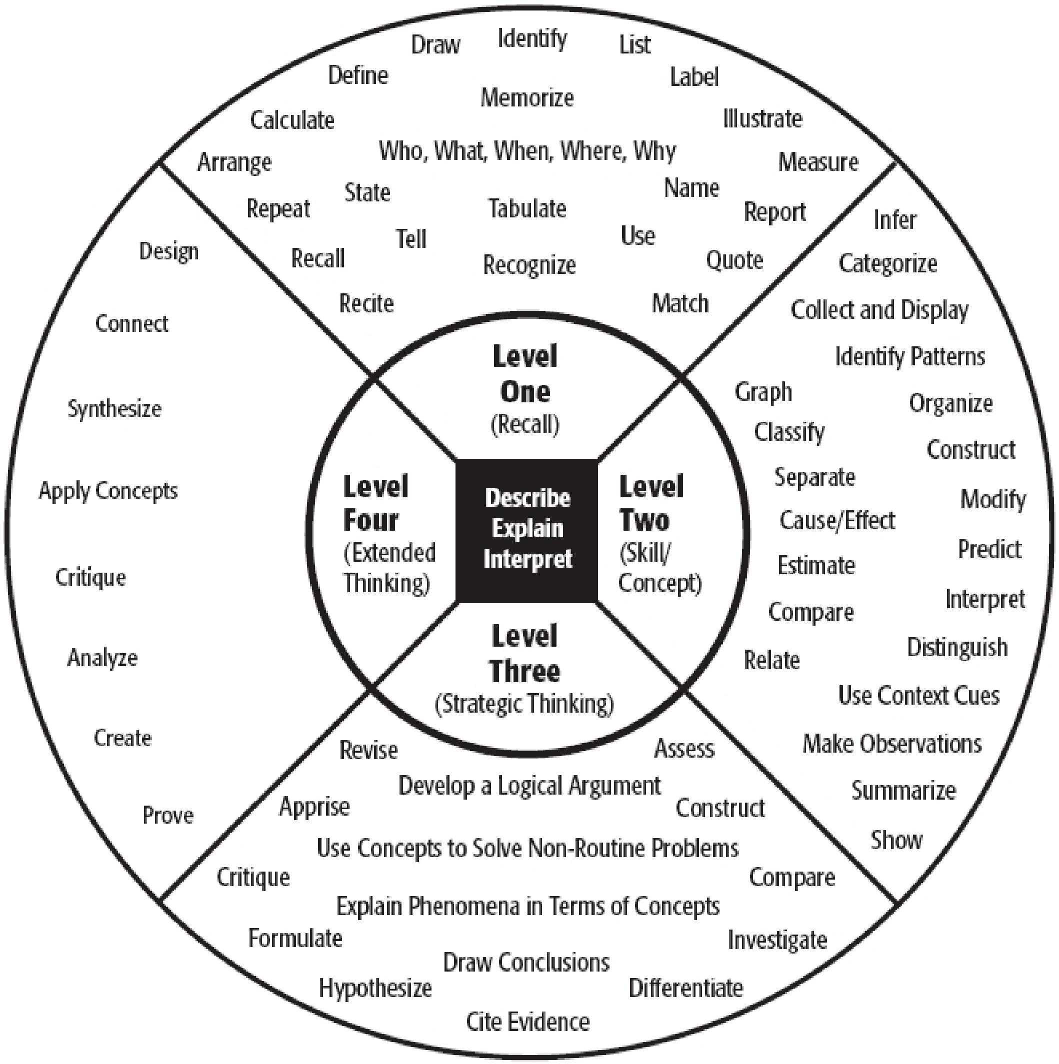
- Higher-order thinking may not lead to deeper understanding. Hess affirms that the vital element required for deeper understanding, transfer, and meaning making, may not always be required within the realm of higher-order thinking. In other words, some prompts or learning exercises may ask the student to apply higher-order thinking skills to complete the task, but the expansion of the student's knowledge and the extrapolation of the students' knowledge and practiced skill to create and discover novel and open-ended solutions may not be required. In essence, the PSAT/NMSQT may ask a test-taker to apply higher-order thinking skills, but the majority of the prompts on the PSAT/NMSQT are close-ended. Resultantly, according to Hess, this lack of opportunity to advance deeper understanding may cause the PSAT/NMSQT to miss the mark as a Career and College Readiness tool (Hess, 2018).
- Undergirding Hess' research in assessing and accurately measuring cognitive rigor is the impression that not all multi-step problems are equivalent. Some educators may believe that a multi-step problem leads to increased levels of cognitive rigor. Yet, a large portion of these multi-step tasks are memorized processes rather than different tiers of understanding. Intrinsically, an educator is obligated to examine the outcome of the

> student's effort and what is the journey the student has to navigate to reach that outcome? Does the learning exercise compel the student to acquire evidence, think critically, look at data through a creative lens? In essence, the length of the process to solve a problem does not equate to the depth of cognitive rigor. Rather, the quality of the process in which the student transfers foundational knowledge and skills and applies that basis to make meaning and extend their thinking to solve a problem (Hess, 2018).

In practice, Hess has applied the Cognitive Rigor Matrix to a multitude of assessments and curricula:

> Although related through their natural ties to the complexity of thought, Bloom's Taxonomy and Webb's Depth of Knowledge (DOK) model differ in scope and application. Bloom's Taxonomy <u>categorizes the cognitive skills required of the brain when faced with a new task</u>, therefore describing the type of thinking processes necessary to answer a question. The DOK model, on the other hand, relates more closely to the depth of content understanding and scope of a learning activity, which <u>manifests in the skills required to complete the task from inception to finale</u> (e.g., planning, researching, drawing conclusions). Today, interpreting and assigning intended DOK levels to both the standards and the related assessment items form critical components of any alignment analysis. Educators have applied Webb's DOK levels across all content areas (Hess, 2004, 2005a, 2005b, 2006a, 2006b; Petit & Hess, 2006). Many states and districts employ DOK to designate the depth and complexity of state standards to align the state's large scale assessments or to revise existing standards to achieve higher cognitive levels for instruction. (Hess et al., 2009)

Figure 10. Webb's DOK Wheel, adapted from *Growing Up Gifted* (1983, p. 222)



The intersection of Bloom's Taxonomy and Webb's Depth of Knowledge is not only the most unique aspect of the Cognitive Rigor Matrix, but the blend of the two frameworks adds to the effectiveness of the HCRM as a tool for educators to deliver instruction and assess students' knowledge:

> Because no simple one-to-one correspondence relates Bloom's Taxonomy and depth of knowledge, Hess (2006b) superimposed them. The resulting cognitive rigor (CR) matrix vividly connects, yet clearly distinguishes, the two schemata, allowing educators to examine the rigor associated with tasks that might seem at first glance comparable in complexity. The cognitive rigor matrix soon found use in states just beginning to appreciate the role cognitive complexity played in test design and item

development. (Hess et al., 2009)

The practical application of the HCRM spans all facets of education, especially in designing a measured approach for the delivery of instruction to learners. The opportunity to reflect upon a prompt, a word problem, a learning task, or a test question can, at times, but a subject task. Analyzing learning experiences through the lens of a subject-specific HCRM is an effective and objective practice to help educators understand on a deeper level what is being asked of their students. Specifically, the single-step math routine resides in the Cell Coordinate on the HCRM [1,3], which translates to Recall on the DOK axis of the matrix and the Understand level on Bloom's axis of the matrix. This type of scoring allows educators and researchers the chance to discover the cognitive rigor of an assessment, a lesson, or overall unit of a curriculum (Hess et al., 2009).

The reliability and the validity of HCRM has stood up to the test of time and a plethora of educational artifacts that support the effectiveness of the tool. For instance, two large-scale studies (The Standards Company LLC, 2008a, 2008b) of mathematics and English language arts curricula analyzed over 200,000 samples of student work, which included homework samples, tests, quizzes, and other assignments and found that, using the HCRM, most language arts assignments correlated to the [2, 2] cell of cognitive rigor, and mathematics assignments corresponded mostly to the [1, 1] and [1, 3] cells. According to Hess and colleagues (2009), these studies indicate that most assignments rely on application of routine steps, rather than higher-order thinking skills.

The integration of Hess' Cognitive Rigor Matrix into a test construction or a curriculum builder takes time and effort:

Does the number of texts read or the complexity of a text always lead to deeper

> understanding and higher DOK levels? Research may say that the task expectations and the level of engagement needed with the text to complete the task are as important as the complexity of the text. We can ask lower level, basic comprehension questions (DOK 1 or DOK 2) using a very complex text. We can ask lower-level, basic comprehension questions using two texts (e.g., compare the settings in these stories). Questions like these would only be DOK 2 because the information needed to answer the question is most likely stated explicitly in the texts. We can ask deeper, more rigorous questions (Becker, 2013). Asking students to draw and support inferences using evidence from illustrations or non print texts can be just as mentally challenging as reading and interpreting words in a print text (Both require justification, so both are DOK 3). (Hess, 2018)

Looking at the educational experience through the lens of a student is integral to the application of the HCRM, as well:

> If we were to ask students to read *Journey and Quest* (the second wordless text in this trilogy by Becker) in order to examine aspects of the author's craft across texts, we set the stage for greater mental challenge and higher cognitive demand (DOK 4). Likewise, when we ask students to use two print texts to analyze the author's craft or the themes in two poems or stories, it becomes a DOK 4 task because the information needed to answer the question is not stated explicitly in the texts and must be interpreted using a deeper level of analysis and broader knowledge base.
>
> Having tasks requiring multiple steps, asking students to read more texts, or using complex data or texts in routine ways doesn't lead to depth of thought. Deeper understanding is achieved when students dive into the concept, the scenario, the model, or the text(s) and come away with new understandings and insights, making connections that transfer to future learning. (Hess, 2018)

On the other hand, Hess' Cognitive Rigor Matrix can be challenging at first glance and takes time for the user to develop accuracy and reliability in determining a learning experience or a test question's cognitive rigor. "Questions and tasks that may be categorized at a lower level along one axis may be considered as a higher level along the other axis. In some instances, it is important to consider what action or behavior follows the verb. Consider the following instance:

• Describe the rule to convert a decimal to a percent.

• Describe an example and a nonexample of a polygon.

• Describe the data represented in the graphs and support statements and/or conclusions.

• Describe the most appropriate dimensions for a box to hold 8 golf balls.

The term described is classified at a lower level (remember) in Bloom's taxonomy. As can be seen in the table, it is more the content of the question or task posed than any particular verb that determines complexity. Likewise, the same argument may apply to "why" and "how" questions to elicit student descriptions and explanations" (Simpson et al., 2014). To that end, Hess' Cognitive Rigor Matrix does have some overlapping wording that detail the traits located in specific cells. The utility of similar language within adjacent cells of an HCRM can be challenging for a coder or a user to distinguish between the classification of the cognitive rigor of a test prompt or a learning exercise. For instance, the reading matrix cell in the Skills & Concepts column and the Understand Row (Coordinate 2,2) states a prompt may ask a student to make basic inferences or logical predictions from data or texts and the adjacent cell in the Strategic Thinking column and the Understand Row (Coordinate 3,2) states a test question may ask a student to make inferences about explicit or implicit themes. The traits are similar, yet different, and to the minimally trained coder, the similar vernacular in some of the cells on the different Hess' Cognitive Matrices may be frustrating or time-consuming to be able to reach a distinct conclusion when reviewing assessment or curricular material.

The Deeper Learning Initiative focuses on enabling students to emerge from their schooling with the ability to master core academic content knowledge, think critically and solve complex problems, work collaboratively, communicate effectively, and learn how to learn. Ideally, a deeper learning assessment would assess elements from each of the cognitive, intrapersonal, and interpersonal domains. However, similar to the state tests analyzed in an earlier study (Yuan & Le, 2012), the tests included in this study were limited in the types of deeper learning skills they measured. Only skills in the cognitive domain were measured, and skills measuring intrapersonal and interpersonal skills were absent or not explicitly assessed. For example, because the tests included in this study were intended as measures of individual students' competencies, none of the tests assessed the ability to work collaboratively. In a similar vein, although students may have engaged in self-reflective learning processes or other "learning how to learn" skills while answering the test items, none of the selected tests explicitly set out to measure learning how to learn. The tests included in this study also measured effective communication in terms of written skills but not in terms of oral skills, so they could measure only limited aspects of effective communication (Yuan & Le, 2012).

Most notably, the distinction of the original Bloom's Taxonomy from Webb's Depth of knowledge lies in the students skills required to access a prompt and the cognitive functions required from inception to completion of a learning exercise. However, Bloom's Taxonomy has been revised since its original conception to be more fluid for the cognitive rigor required of 21st century learners. Specifically,

> the revised descriptors [of Bloom's Taxonomy] consider both the processes (the verbs) and the knowledge (the nouns) used to articulate educational objectives. This restructuring of the original taxonomy recognizes the importance of the interaction between content (characterized by factual, conceptual, procedural, and metacognitive knowledge) and thought processes. However, educators in the 21st century still yearn for cognitive rigor rather than a list of verbs that scale upward as the taxonomy yields higher level thinking descriptors. Essentially, Bloom's taxonomy provides a solid foundation for teachers to introduce students to skills and content that require higher order thinking skills, but when the educator seeks to formatively or summatively assess a student's skills, Bloom's taxonomy does not quite measure up to a universal depth of understanding educators are looking for in an effective assessment of their students. (Hess et al., 2009)

The *Taxonomy of Educational Objectives* (1954) is a framework for classifying statements of what we expect or intend students to learn as a result of instruction. The framework was conceived as a means of facilitating the exchange of test items among faculty at various universities in order to create a bank of items, each measuring the same educational objective. Bloom, then Associate Director of the Board of Examinations of the University of Chicago, initiated the idea, hoping that it would reduce the labor of preparing annual comprehensive examinations. The original taxonomy provided carefully developed definitions for each of the six major categories in the cognitive domain. The categories were: *Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation.* With the exception of *Application*, each of these was broken into subcategories. The categories were ordered from simple to complex and from concrete to abstract. Further, it was assumed that the original taxonomy represented a cumulative hierarchy; that mastery of each simpler category was a prerequisite to mastery of the next more complex one. The structure of the taxonomy cascades into tiers of objectives that describe learning objectives as a result of the teacher's instruction or what is being asked of a student. The original version of Bloom's taxonomy tiered words that were not reasonably sufficient in illustrating the learning objectives of a prospective educational environment. Subsequently, the revisions were focused on the functionality and operationalization of the words to be applied to a dynamic, engaging, reflective, and active learning environment. The journey of the revisions landed upon six main categories (remembering, understanding, applying, analyzing, evaluating, and creating) that were carefully crafted to ensure users were able to prioritize the learner's metacognition (Conklin, 2005). As readers saw its potential, the framework became widely known and cited, eventually being translated into 22 languages.

Figure 11. Revised Bloom's Taxonomy adapted from Anderson and Krathwohl (2001)

Revised Bloom's Taxonomy – Categories and Cognitive Processes					
Lower order thinking skills		→			**Higher order thinking skills**
Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Recognizing (identifying) Recalling (retrieving)	Interpreting (clarifying, paraphrasing, representing, translating) Exemplifying (illustrating, instantiating) Classifying (categorizing, subsuming) Summarizing (abstracting, generalizing) Inferring (concluding, extrapolating, interpolating, predicting) Comparing (contrasting, mapping, matching) Explaining (constructing models)	Executing (carrying out) Implementing (using)	Differentiating (discriminating, distinguishing, focusing, selecting) Organizing (finding coherence, integrating, outlining, parsing, structuring) Attributing (deconstructing)	Checking (coordinating, detecting, monitoring, testing) Critiquing (judging)	Generating (hypothesizing) Planning (designing) Producing (construct)

In essence, "the authors of the revised taxonomy underscore this dynamism, using verbs and gerunds to label their categories and subcategories (rather than the nouns of the original taxonomy). These "action words" describe the cognitive processes by which thinkers encounter and work with knowledge. A statement of a learning objective contains a verb (an action) and an object (usually a noun). The verb generally refers to [actions associated with] the intended cognitive process. The object generally describes the knowledge students are expected to acquire or construct (Anderson & Krathwohl, 2001, pp. 4–5). The cognitive process dimension represents a continuum of increasing cognitive complexity—from remember to create. Anderson and Krathwohl identify 19 specific cognitive processes that further clarify the bounds of the six categories (Iowa State University Center for Excellence in Learning and Teaching, 2020).

Most importantly, the updated design of Bloom’s Taxonomy enables educators “to develop good questions with Bloom's Revised Taxonomy, [so that they] can use the basic categories from Bloom's Questioning Inverted Pyramid to replace the cognitive verbs of performance objectives with the following question stems:

Recognize: Who? What? Where? When?

Understand: How? Why?

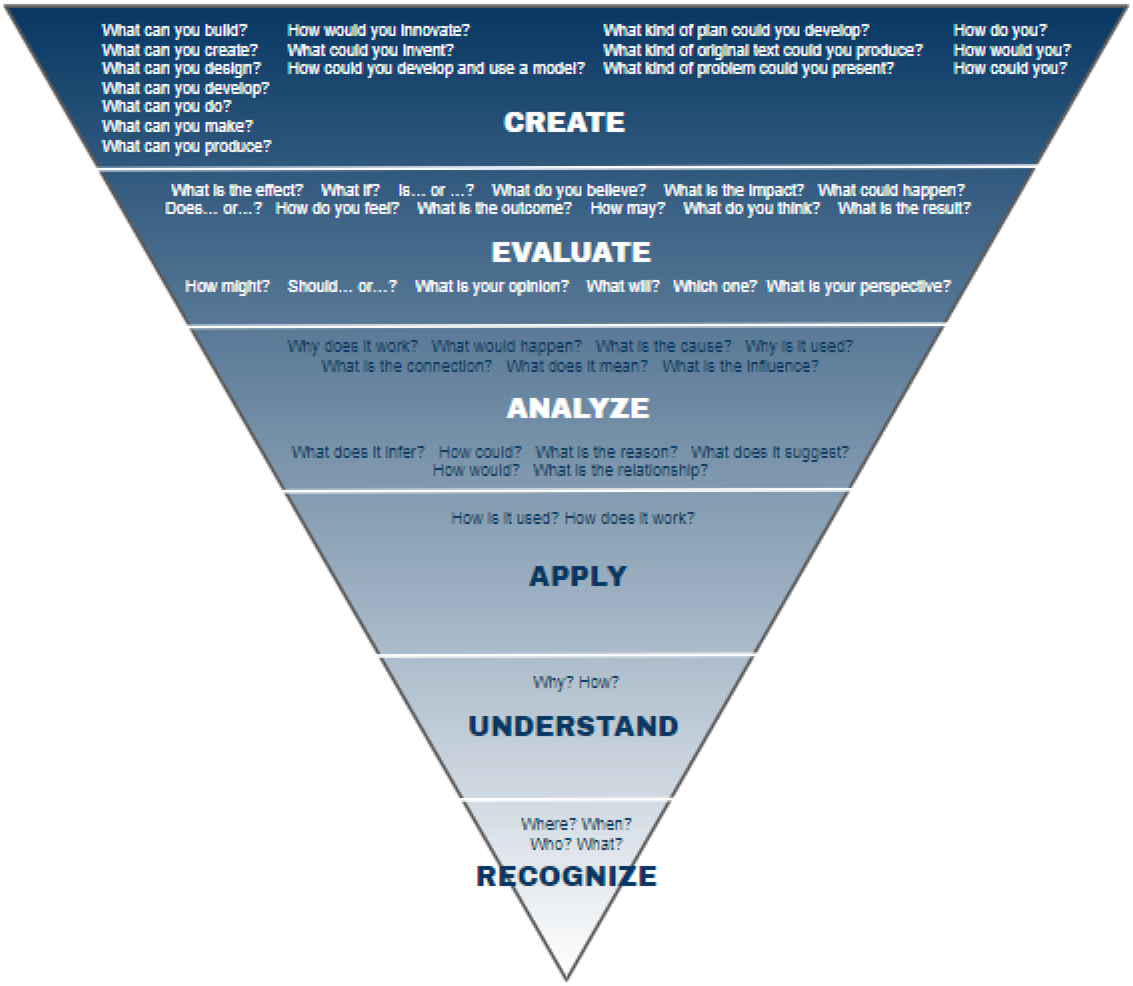
Apply: How does it work? How is it used?

Analyze: Why does it work? Why is it used? What does it imply/infer? What does it mean? What does it suggest? What is the cause? What is the connection? What is the influence? What is the reason? What is the relationship?

Evaluate: What is the effect? What is the impact? What is the outcome? What is the result? What if? What would happen? What could happen? What do you believe about it? How do you feel about it? What do you think about it? What is your opinion about it?

Create: What can you create? What can you design? What can you develop? What can you plan? What can you produce? How could you innovate? What could you invent? How do you? How could you? How would you? How could you develop and use a model? What kind of original text could you produce? What kind of problem could you present” (Francis, 2016)?

Figure 12. Inverted Triangle of Revised Bloom's Taxonomy adapted from Anderson & Krathwohl (2001)



The inverted triangle diagram encapsulates and illustrates the nature of the higher-order thinking skills students need to apply to complete a learning task properly. The Revised Bloom's Taxonomy is another effective tool that has withstood the test of time to accurately assess a student's academic growth.

The categories in <u>Taxonomy of Educational Objectives: Cognitive Domain</u>, knowledge, comprehension, application, analysis, synthesis, and evaluation are the major categories of processes. The essential structure of the Taxonomy was a cumulative hierarchy: hierarchy because the classes of objectives were arranged in order of increasing complexity, and cumulative because each class of behaviors was presumed to include all the behaviors of the less complex classes (Kreitzer & Madaus, 1994).

In the years following the publication of the taxonomy, critics and researchers generated philosophical and empirical evidence against the claim that the levels in the taxonomy are hierarchical (Anderson & Sosniak, 1994). In this hierarchy, knowledge and comprehension were regarded as lower-level processes, whereas application, analysis, synthesis, and evaluation were considered higher-level processes.. Additionally, criticisms focused on distinctions made among categories cited in his work that could not be defined clearly and were open to some interpretations (Bloom et al., 1956). Indeed, in response to the critics, Bloom wrote, "It is obvious at least to me, that many of the criticisms directed toward the taxonomy have resulted from very narrow interpretations of both the taxonomy and its proper application" (Anderson, 2020). According to Kastberg (2003), the taxonomy is used in many parts of the world as a useful tool for teachers as they create their assessments.

Similar Studies

Specific research on the PSAT/NMSQT is limited in terms of the language used in each question as compared to Hess' Cognitive Rigor Matrix definition of higher-order thinking; however, similar studies were performed to identify the link between test scores and student performance. A few parallel studies looking at standardized test performance in relation to managerial or job-performance success found a link between aptitude and occupational output:

> For example, Greener and Osburn (1979) examined the ability–performance relationship of 5,900 white-collar managerial and professional employees using the Miller Analogies Test, Non-Verbal Reasoning Ability Test, Guilford-Zimmerman Temperament Survey, a company-developed survey of biographical information, and an overall employee effectiveness rating to predict managerial success. They found a linear relationship

> between the ability measures and managerial success. Reviewing the literature on the General Aptitude Test Battery (GATB)–job performance relationship, Hawk (1970) found that while some of the 367 studies showed a nonlinear relationship existed between the GATB and job performance, this number did not exceed the number expected by chance, thus indicating a linear relationship exists. Similarly, based on 174 studies (N = 36,614), Coward and Sackett (1990) conducted a meta-analysis to examine whether the relationship between the nine scales of the GATB and job performance is linear. Overall, they found that the power polynomial approach had more statistical power in detecting nonlinearity. (Marini, Mattern, & Shaw, 2011)

Some of the more exacting studies in relation to the validity of the PSAT/NMSQT and the standardized assessments of the educational field that require higher-order thinking skills and purport to indicate college and career readiness have been conducted in an attempt to tilt the idea that standardized assessments are basically filters for ridding evaluators of students who do not perform well. Efforts have been made to illustrate the relationship between the scores of students on standardized test scores and other academic outcomes. Specifically, based on nearly 50,000 students across 13 universities, a study by Sackett, Hardison, and Cullen (2004) found that the relationship between SAT scores and college GPA by gender and race was linear.

Similarly, Arneson, Sackett, and Beatty (2011) investigated the linearity of the ability–performance relationship among high-scoring students. Three different datasets were analyzed, including data from the College Board with SAT verbal (now critical reading) and math scores and first-year GPA (FYGPA) for about 167,000 students; data from the National Educational Longitudinal study of 1988 (NELS:88) with self-reported grades for approximately 25,000 eighth-graders; and data from Project TALENT with self-reported HSGPA for about 300,000 high school students (grade 9–12). Based on the entire score scale, the results indicated a linear relationship between SAT scores and FYGPA since the addition of a quadratic term (i.e., SAT2) did not result in a statistically significant change in model fit. However, for the NELS:88

and Project TALENT datasets, there was slight evidence of a deviation from linearity for the ability–grades relationship across the entire score scale.

Arneson and colleagues' 2011 study, which also examined the linearity of the ability–performance relationship among high-scoring students, is particularly relevant to the current study. In the upper tail of the SAT-FYGPA relationship, defined as one standard deviation above the mean SAT score and higher (*M* = 1140, *SD* = 169), slight deviations from linearity were detected. It should be pointed out that even though the quadratic term was statistically significant, the effect size was small, with a ΔR2 of 0.000. In the NELS:88 and Project TALENT datasets, the top of the ability distributions did not deviate from linearity (Arneson, Sackett, & Beatty, 2011).

Finally, another level of investigation research on different types of students in a longitudinal sense. A study was conducted to reveal the long-term effects on the academic outputs and the application of higher-order thinking skills of high scoring students across a few decades:

> Research on gifted children can also shed light on the question of whether test scores meaningfully discriminate among very high-scoring students. Research started by Keating and Stanley (1972) investigated gifted students' performance on the SAT as part of the Study of Mathematically Precocious Youth (SMPY). They collected data from seventh- and eighth-graders who qualified as being gifted and were given the opportunity to take the SAT during their seventh- or eighth-grade year. Benbow (1992) analyzed the top 1 percent of the scorers in the first SMPY cohort (1972–1974) and found that the top quartile of the top 1 percent earned significantly higher college GPAs compared to those of the bottom quartile of the top 1 percent. Statistically significant results were also found for variables such as intellectual level of college attended and math/science GPA among other results. Wai, Lubinski, and Benbow (2005) extended Benbow's (1992) study by using the SMPY data to examine the relationship between SAT scores and occupational outcome measures. The first two cohorts of SMPY data, which included data collected 20 years after the original collection, were analyzed. Similar to the analyses conducted by Benbow (1992), the top and bottom quartile of the top 1 percent of scorers in seventh and eighth grades were compared in relation to earned doctorates, primary income, number of patents, and tenure 20 years later. A significant

> difference was found for each of the variables, indicating that not only can comparisons be made in such a selective sample, but also that scores from adolescence are predictive of occupational outcomes 20 years later. (Marini, Mattern, & Shaw, 2011)

Students learn skills and acquire knowledge more readily when they can transfer their learning to new or more complex situations, a process more likely to occur once they have developed a deep understanding of content (Pellegrino, 2012). Therefore, ensuring that a curriculum aligned to standards alone will not prepare students for the challenges of the twenty-first century. Therefore, teachers must provide all students with challenging tasks and demanding goals, structure learning so that students can reach high goals, and enhance both surface and deep learning of content (Marsh & Hattie, 2002). Hence, both Bloom's Taxonomy and Webb's depth of knowledge serve important functions in education reform at the state level in terms of standards development and assessment alignment. Because cognitive rigor encompasses the complexity of content, the cognitive engagement with that content, and the scope of the planned learning activities, the HCRM can enhance instructional and assessment practices at the classroom level as well. Superposing the two cognitive complexity measures produces a means of analyzing the emphasis placed on each matrix intersection. As educators become more skilled at recognizing cognitive rigor and analyzing its implications for instruction and assessment, they can enhance learning opportunities for all students and across all subject areas and grade levels. Because students need exposure to novel and complex activities every day, schools in the twenty-first century should prepare students by providing them with a curriculum that spans a wide range of the cognitive rigor matrix (Hess et al., 2009).

Chapter III

METHODOLOGY

Introduction

This qualitative document analysis focused on comparing, synthesizing, and describing the language of complex thinking embedded within the PSAT/NMSQT. The ever-expanding prominence of the claim that the PSAT/NMSQT is an assessment of college readiness needs to be investigated for several compelling reasons. The importance of the PSAT/NMSQT has increased over the course of time to propel students on different paths and academic trajectories. Specifically, a student in New Jersey will have their score used to indicate proficiency and eligibility for high school graduation. Additionally, the College Board refers to the PSAT/NMSQT score to recommend students take a more challenging Advanced Placement course that may allow a student to earn college credit (thus saving on the cost of college and having the most rigorous transcript possible). Lastly, the PSAT/NMSQT can also be used for placement into high school honors levels or as an obstacle preventing students from entering honors-level courses.

Essentially, the PSAT/NMSQT has transcended from a formative assessment to give students a practice attempt at the SAT to now splintered into a tracking device and a gatekeeper for students. To illustrate, a *Forbes* article focused on the importance of the PSAT detailed how:

> High school guidance counselors are typically responsible for about 500 students each (for public schools, the national average is 475 to 1). This astounding ratio means that many don't have time, at least initially, to think of individual students as much more than a GPA, a transcript, and a test score. A score on the PSAT that is significantly lower than might be expected based on GPA or transcript may, therefore, affect the initial advice that an overburdened counselor gives a student" (Edmonds, 2015). By the same token, *The Huffington Post* highlighted the power of the PSAT, and illustrated that the PSAT offers a college admissions edge. There are some 36,000 high schools in America, about

> 26,000 public schools and about 10,000 private schools. That means 36,000 valedictorians. Multiples more than that of 4.0 (unweighted) or close GPA students. That means multiple tens of thousands of football, basketball, softball, soccer, volleyball, debate, and math team captains. But high schools are not all the same and with so many, college admissions people can only become familiar with the reputation of a very small number. That's why national standardized tests and national level awards and recognitions matter so much. It proves a student is not just the biggest frog in a small pond, but truly a standout in a huge national ocean. (Peterson, 2015)

The article goes on to highlight how the PSAT has far-reaching implications. Peterson (2015) noted that only the PSAT taken in the students' junior year counts, and that students have one shot to succeed, leading to massive stress on the part of students and their parents.

Accordingly, the path of this examination of the PSAT/NMSQT assessment will scrutinize the language used in each question of each section in the 2018 PSAT/NMSQT through the lens of Hess' Cognitive Rigor Matrix (HCRM) tool. The HCRM is a research-based tool that will be used to identify what each question is asking in terms of critical 21st-century career and college readiness skills. Invariably, the research questions have been crafted to focus on the language of the questions throughout the PSAT/NMSQT, and what the student or the test-taker is asked to do in terms of higher-order thinking as they engage each component of the college-readiness assessment. The researchers believe the PSAT/NMSQT as compared to the HCRM will be indicative of the commitment the College Board's test framers made to propelling students to success. In essence, communication, collaboration, critical thinking, and creativity have been identified as crucial to the success of employees, employers, and contributors to democracy living in the Information Age (Dintersmith, 2015). Specifically, higher-order thinking skills such as transfer and meaning making are almost seen as invaluable currency in today's entrepreneurial, data driven economic arena. As a result, this study of the PSAT/NMSQT should reveal the influence the Information Age's definition of required higher-order thinking skills has on the PSAT/NMSQT's claim they are supposedly evaluating the career and college-readiness of high school students.

Research Questions

1. In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Reading Section questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?
2. In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Math section (both calculator and non-calculator) questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?
3. In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Writing section's questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?

Research Design

The priority of the research plan is to better understand the language contained in each question on the 2018 PSAT/NMSQT and what higher-order thinking skills that question's language requires of the test-taker. The qualitative document analysis (Corbin & Strauss, 2008) was chosen to align with several other similar studies:

> A document analysis is a systematic procedure for reviewing or evaluating documents-both printed and electronic (computer based and Internet-transmitted) material. Like other analytical methods in qualitative research, document analysis requires that data be examined and interpreted in order to elicit meaning, gain understanding, develop empirical knowledge (Corbin & Strauss, 2008).

In closely reviewing the language of each question on the PSAT and comparing the language to the framework of the HCRM, the researchers have a goal to "seek convergence and coorbation through the use of different data sources and methods" (Bowen, 2009). Therefore, "by triangulating data, the researcher attempts to provide a confluence of evidence that breeds credibility" (Eisner, 1991).

Subsequently, Hess' Cognitive Rigor Matrix (HCRM) was consciously chosen to support the investigation of the answers of the research questions because the HCRM assists teachers in applying what cognitive rigor might look like in the classroom and guides test developers in designing test items and performance tasks. Content-specific descriptors in each of the HCRMs are used to categorize and plan for various levels of abstraction – meaning an analysis of the mental processing required of assessment questions and learning tasks (Hess, 2014).

To underscore the relevance and the effectiveness of using the HCRM as the tool to analyze the PSAT/NMSQT, Bowen details the foundation of a document analysis of which "documents provide supplementary research data. Information and insights derived from documents can be valuable additions to a knowledge base" (Bowen, 2009). As a result, the measurement of each question as provided by the HCRM's scale enables the researcher to gain a thematic understanding of the PSAT/NMSQT and then clearly display the findings of the document analysis. Quite simply, the PSAT/NMSQT claims to be "a meaningful step of a student's path to college" (College Board, 2021) and, with the HCRM being an authentic, well-researched tool that is designed to analyze assessment questions, the researchers have triangulated the data to measure and interpret the credibility of how meaningful the PSAT/NMSQT evaluates a student's college and career readiness.

Data Sources and Collection

In circling back to the research questions, the study will investigate the crafted language of each question of each section of the 2018 PSAT/NMSQT through the lens of the HCRM. The acquisition of sample questions is a fairly straightforward process. After the scores are released to the students (typically in early December of the testing year), the test questions and the test booklets become available to the public by the College Board. The 2018 PSAT/NMSQT Test Book is the primary document that will be studied and supplemented by the HCRM to triangulate the data. The analysis of the document is based upon a careful, more focused re-reading and review of the test questions. The researchers will be taking a closer look at the characteristics of all of the questions of the 2018 PSAT/NMSQT to uncover what higher-order thinking skills are interwoven throughout the test. The 2018 PSAT/NMSQT contains 41 English Language Arts Questions, 33 Math Questions (divided into calculator and a non-calculator subsections), and 39 writing questions (that includes multiple-choice and open-ended responses). The researchers meticulously paced every single question of the 2018 PSAT/NMSQT through the HCRM to be as thorough as possible. Each question's score will be placed into a spreadsheet that contains the question number, the test section, the type of question, the score on Bloom's portion of the HCRM, and the score on Webb's portion of the HCRM. Data from the spreadsheet will be sorted based upon section, rigor, and frequency of the score. Visualizing the data in terms of frequency, range, mean, median, and mode offered researchers a picture of how the language used on the PSAT/NMSQT assesses students' higher-order thinking skills. Explicitly the purpose of each statistical analysis has an essential purpose in illustrating how the the language of each question of each section of the PSAT/NMSQT compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking.

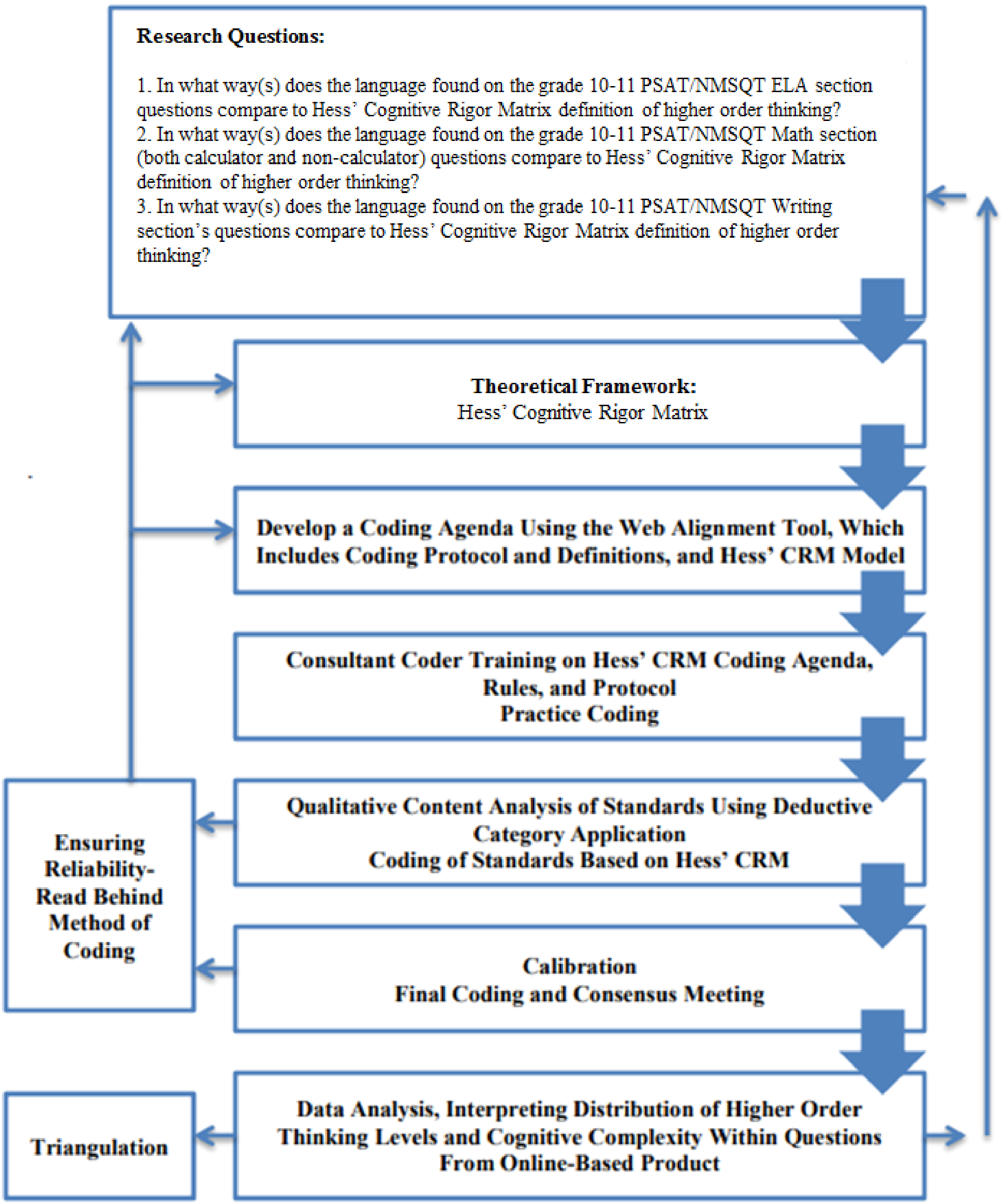
- Frequency - identifying of the number of each score each question earns depicts the distribution of the level of cognitive rigor across the 2018 PSAT/NMSQT
- Range - illustrating the highest level of cognitive rigor as scored by the HCRM and the lowest level of cognitive rigor scored by the HCRM provides the scope of cognitive rigor required for the test-taker on the 2018 PSAT/NMSQT.
- Mean - calculating the mean for the overall scores of the 2018 PSAT/NMSQT and for each section of the PSAT/NMSQT finds the center of the scores in the data collection.
- Median - calculating the median for the overall scores of the 2018 PSAT/NMSQT and for each section of the PSAT/NMSQT finds the center of the scores in the data collection that can better account for and reveal if the mean is skewed higher or lower.
- Mode - calculating the mode for the overall scores of the 2018 PSAT/NMSQT and for each section of the PSAT/NMSQT displays the higher-order thinking skills and cognitive rigor skills that are used the most overall and for each section.
- Percentage of HCRM score for each PSAT/NMSQT section - utilizing percentages is another illustration of the distribution of higher-order thinking skills and cognitive rigor for each section

Most importantly, Hess' Cognitive Rigor Matrix (HCRM) was intentionally chosen as the tool to evaluate the cognitive rigor of the 2018 PSAT/NMSQT because of the adaptability and alignment the HCRM directly has to each section of the PSAT/NMSQT. Unique to each PSAT/NMSQT section evaluation is a prefabricated and specifically designed HCRM. The

Reading Section of the PSAT/NMSQT has a Close-Reading & Listening Cognitive Rigor Matrix, the Math Section of the PSAT/NMSQT has a Math & Science Cognitive Rigor Matrix, and the Writing and Language Section of the PSAT/NMSQT has a Written & Oral Communication Cognitive Rigor Matrix.

Notably, the HCRM provides predefined codes to generate data for the analysis portion of the study in Chapter IV. Emphatically, the researchers have upheld objectivity as they seek to represent the research material fairly and with sensitivity (responding to even subtle cues to meaning in the text of the questions) (Bowen, 2009). Moreover, deductive category application is the process in which text is analyzed based on pre-existing categories following a coding protocol (Marying, 2000). The pre-existing categories for the study were the categorization of thinking as represented in the HCRM.

Figure 13. Step model for deductive category application adapted from Mayring (2000) and Sydoruk (2018).



Additionally, the researchers infused Webb's Alignment Training Manual that included Tips for Facilitating the Consensus Process. The following facilitator tips were used during the coding process:

1. Read each test question aloud before discussing it.

2. As you go through the test questions, actively solicit comments from all reviewers.

3. Use your printout to call on people who coded DOK differently from the coding of other members of the group, and ask them to explain why they coded the test question to the particular DOK level.

4. Once two reviewers have described how they have coded a test question differently, ask a third reviewer to highlight the differences between the two interpretations.

5. Restate and summarize to reviewers your interpretation of what the reviewers have agreed on and what they have disagreed on.

6. If there is a difference in interpretation of the test question's terminology to a reviewer with experience in teaching that grade level with these standards, discern how the teacher might be interpreting the test question.

7. Ask if anyone, through other reviewers' explanations, now want to change his or her mind about their original coding.

8. If the viewpoints on the DOK level of a test question is divided, point to the most likely skills or content knowledge required in the test question, not the more extreme possibilities the test question might allow for.

9. As the facilitator, try not to dominate the consensus process. Even if you have strong *feelings about the DOK level of a test question, wait to see if other reviewers highlight your point (Webb et al., 2005).*

Lastly, the structure and the focus of the qualitative, deductive coding analysis of each section of the PSAT/NMSQT does not involve individual participants. The study is using a document as the crux of the research. Therefore, protective measures are not required for any "participants."

Data Analysis

As a result of the qualitative, deductive coding analysis, the scores were calculated, analyzed, summarized, and have been illustrated in a detailed synthesis of each section of the PSAT/NMSQT through a series of tables. Chapter IV contains all of the tables that show distinct trends and findings evident through the lens of the research questions in exploring the language used on the questions of the PSAT/NMSQT and ultimately the access and application of higher-order thinking skills. The goal of each data table is to indicate the level of depth and the meticulous nature of the study in which each question of each section of PSAT/NMSQT testing was analyzed and scored. The numeric value that is displayed in the data table is the aggregate result of how each question was scored based upon the scored level on the HCRM. Specifically, Hess' Cognitive Rigor Matrix for English Language Arts (HCRM-ELA) was used to score the Reading Section of the PSAT/NMSQT reached in concordance with the HCRM tool that was used to classify each standard. The descriptors and keywords contained in HCRM show how the wording of the test questions on the PSAT/NMSQT can align to each of the Levels, thus providing a path for the researchers to illustrate how the language found on the grade 10-11 PSAT/NMSQT compares to Hess' Cognitive Rigor Matrix definition of higher-order thinking.

The HCRM provided the researchers with an authenticated, researched, and tested-over-time instrument to accurately evaluate the rigor of the PSAT/NMSQT as compared to the rigor matrix detailing the specificity of rigor asked of the students on the assessment. Notably, the rows and columns of the HCRM corroborate Bloom's Taxonomy and Webb's Depth of Knowledge to provide the evaluator with an effective, concise, and precise tool to measure the rigor of a test question. A sample of the matrix scoring model once the test question is evaluated can be seen in the table below:

Table 3. Scoring Matrix adapted from Hess' Cognitive Rigor Matrix (Sydoruk, 2018)

Webb's → **Bloom's ↓**	**Recall and Reproduction [1 ,_]**	**Skills and Concepts [2 , _]**	**Strategic Thinking [3 , _]**	**Extended Thinking [4 , _]**
Remember [_ , 1]	[1,1]	**N/A**	**N/A**	**N/A**
Understand [_ , 2]	[1,2]	[2,2]	[3,2]	[4,2]
Apply [_ , 3]	[1,3]	[2,3]	[3,3]	[4,3]
Analyze [_ , 4]	[1,4]	[2,4]	[3,4]	[4,4]
Evaluate [_ , 5]	**N/A**	**N/A**	[3,5]	[4,5]
Create [_ , 6]	[1,6]	[2,6]	[3,6]	[4,6]

As one can see from *Table 3* above, an analyzed test question was then assigned a score, and placed into a spreadsheet. Succinctly, the questions were scored and tabulated into columns based Webb's Depth of Knowledge column, with Bloom's Taxonomy, and the bar graph shows the distribution of the scoring system using the matrix cross-sectional score (i.e, [1,2], [2,3], [3,2], etc.). Once all of the test questions are analyzed and the spreadsheet contains all of the individualized scores for each question, the data will be sorted, synthesized, and displayed for others to conclude in Chapter IV. Invariably, Chapter IV will display a summary of the correlational statistical analysis to provide insight into the effectiveness of the PSAT/NMSQT to measure higher-order thinking skills.

Validity and Credibility

In reviewing the basis of the research questions, the study will investigate the PSAT/NMSQT through the lens of the HCRM. Significantly, a dedication to reliability and validity will be made throughout the data analysis process. The reliability of a study can be summarized as the process in which the study can be repeated and yield the same results. Also, the data analysis structure replicated that of previous studies that used Webb's Depth of Knowledge to qualitatively analyze an educational text such as curriculum standards. Additionally, this study involved two analysts coding each of the PSAT/NMSQT questions then comparing their data and findings, thus increasing inter-rater reliability. Accordingly, in this study, the analysts used the same coding agenda rules of coding and data as Fitzhugh's (2018) study and a similar structure to Niebling's (2012) study regarding Determining the Cognitive Complexity of the Iowa Core in Literacy and Mathematics in 2012.

Another layer to increase the reliability of the study was to apply a double-rater read-behind model (Fitzhugh, 2018).

> Briefly, this model calls for one rater to independently assign DOK codes to the standards, while the second rater reviews the codes of the first rater to determine if he/she agrees, noting agreement and disagreement. The raters discussed any

discrepancies in an ongoing manner, working to achieve consensus on those discrepancies (Niebling, 2012).

Together with Sforza's (2014) study on the alignment of Common Core State Standards and New Jersey Core Curriculum Content Standards to Webb's Depth of Knowledge and Sydoruk's (2018) study of a Grade 8 Online-Based English Language Arts Skills Program, Hess' Cognitive Rigor Matrix was used as a means of triangulation. Sforza's methods of using a second coder and calibration sessions to ensure that the researchers agreed on the placement of questions into cells within Hess' Cognitive Rigor Matrix were applied to the current study. In addition, Sforza's data collection used Webb's Alignment Tool as a means of reaching agreement into the placement of standards into Webb's Depth of Knowledge, providing information on ways to utilize the Alignment Tool for this study (Sydoruk, 2018). Likewise, the researchers were properly trained to use HCRM application, as delineated in *A Local Assessment Toolkit to Promote Deeper Learning: Transforming Research Into Practice* (Hess, 2018).

Accordingly, validity within a study is the accuracy in which the instrument measures the concept of interest. In this case, the concept of interest is higher-order thinking skills. Therefore, the guidelines in analyzing the rigor of an assessment provided by *A Local Assessment Toolkit to Promote Deeper Learning: Transforming Research Into Practice* were followed. To elaborate, a conscious decision was made to ensure a common expectation of analysis for each standard was met to ensure validity. A partnership was formed between the two researchers to consistently meet in person once a week to train on the HCRM tool for about eight weeks before the analysis of the 2018 PSAT/NMSQT. The researchers worked together in grappling with the evaluation process of the PSAT/NMSQT by examining the 2019 version of the assessment to increase their proficiency. After the researchers established an acceptable agreement rate of 85% on the rigor of the PSAT/NMSQT by examining the 2019 version using the HCRM tool, the researchers moved on to comparing the PSAT/NMSQT 2018 version. The amount of time the researchers dedicated to practicing using the HCRM tool aligns with other studies that have been used to measure curriculum standards and assessments. If any disagreements occurred in the evaluation of a test question's scoring using the HCRM, the two coders consulted with one another to resolve the differing viewpoints, and arrive at a mutually agreed-upon score. After all of the questions of the 2018 PSAT/NMSQT were deductively coded using the HCRM, the results of this analysis were compared to focus on how the language compared to the HCRM definition of higher-order thinking skills.

Limitations

Crafting a thorough document analysis study has advantages and disadvantages. The research method was chosen because of the efficient analysis of the documents, the availability of the PSAT/NMSQT test questions, the stability and lack of malleability of the test questions as opposed to an interview subject, the exactness of the test questions standing the test of time, and the purpose of the test question provides coverage of a time period in which college and career readiness are at the forefront of public education. However, the research methodology chosen does indeed have limits as well. The document analysis has been limited to one year's worth of PSAT/NMSQT questions. The totality of the document analysis is thorough, but the 2018 PSAT/NMSQT is one issue of the running series of the PSAT/NMSQT catalog. Without qualitatively studying the creation and the impressions of the creators of the test questions may leave some readers wanting more of the rationale behind the construction of the test or more of an understanding of how the question was created and for what purpose (what skill did the test question creator look to assess?). In essence, the static nature of a document containing text may not reveal the nuance, the background, or the design process of the final output of the College Board's 2018 version of the PSAT/NMSQT. In fact, Bowen warns of this approach stating:

> Documents should not be treated as necessarily precise, accurate, or complete recordings of events that have occurred. Researchers should not simply 'lift' words and passages from available documents to be thrown into their research report. Rather, they should establish the meaning of the document and its contribution to the issues being explored. (Bowen, 2009)

Lastly, the exhaustive training regimen to increase interrater reliability and reduce bias in the eight weeks leading to the analysis of the 2018 PSAT/NMSQT can be subjected to scrutiny. Another group of researchers may put in more time to practice before the rating of the document using the HCRM, and another group may use more than two coders to make conclusions about the cognitive rigor of a test question.

Summary

This chapter detailed the methods and approaches the researchers have taken in looking into the depth of skills accessed by test-takers for the 2018 PSAT/NMSQT.

> A qualitative content analysis research methodology was applied to answer the three research questions. Furthermore, Mayring's (2000) step model for deductive category application was used to create a visual representation of the research process, including methods to ensure credibility in the overall study. (Sydoruk, 2018)

The basis of the analysis was formulated in applying Hess' Cognitive Rigor Matrices (English Language Arts, Mathematics, and Writing versions). In a parallel process, the critical nature of having two trained coders in which specific criteria were outlined for the coders to identify the level of rigor for each question of the PSAT/NMSQT. If consensus was not reached, in that each coder chose a different level for an evaluation of a question, the coders agreed to choose the higher of the two levels, in accordance with Webb et al. (2005).

Chapter IV will illustrate the data findings of the research with a focus on answering each of the three research questions.

Chapter IV

RESULTS

Introduction and Overall Findings

Chapter IV consists of the data acquired from utilizing Hess Cognitive Rigor Matrices (HCRM) to evaluate every question on the 2018 PSAT/NMSQT. In total, each of the 139 questions was reviewed and then compared against the language requirement of the HCRM that aligned to each of the sections. The Math sections of the PSAT/NMSQT (both the calculator and the non-calculator section) were scored using the Math and Science Cognitive Rigor Matrix. The Reading section of the PSAT/NMSQT was scored using the Close Reading and Listening Cognitive Rigor Matrix. The Writing section of the PSAT/NMSQT was scored using the Written and Oral Communication Rigor Matrix. Notably, the appendix contains the HCRMs that the coders used to evaluate each question. The coders, who had been trained to properly apply the standards and traits of the HCRM, met a total of five times to study, discuss, grapple with the scoring, and build consensus for each of the questions on the PSAT/NMSQT. In alignment with similar studies using the HCRM, the coders agreed that any question scoring a 3 or 4 on Webb's Depth of Knowledge scale would be considered a higher-order thinking question. In the scoring matrix, higher-order thinking scores include: [3,2], [3,3], [3,4], [3,5], [3,6], [4,2], [4,3], [4,4], [4,5] and [4,6]. In duplicating Sydoruk's study (2018) and Fitzhugh's study (2019) in tabulating questions using the HCRM, the two coders used the double-rater, read-behind consensus model to reach alignment of each of the questions of the 2018 PSAT/NMSQT (Sydoruk, 2018), allowing both coders to discussion each question's placement in a way to increases inter-rater reliability. In essence, during each of the scoring discussions, the two coders spent time methodically working through each of the questions on the PSAT/NMSQT and reached an agreement upon the final scoring. Once the task of scoring each question was finished, the researchers tabulated the data, organized the data into charts, bar graphs, and heat maps (Zenko, 2021). The researchers chose to illustrate the data of the study using a heat map design to offer the reader a contextualized comparison of the frequency of each cell's language being used by the 2018 PSAT/NMSQT.

The qualitative document analysis was focused upon the nuanced language contained within each question of the PSAT/NMSQT. Regarding the question of how does the language found on the grade 10-11 PSAT/NMSQT questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking, the researchers arrived at a clear conclusion. Overall, 44 out of the 139 questions (31.6%) on the 2018 PSAT/NMSQT contained language as compared to the specified HCRMs definition of higher-order thinking. The data showed 31.6% of the questions required the higher-order thinking skills of Strategic Thinking and Reasoning (DOK Level 3) or Extended Thinking (DOK Level 4). Subsequently, 95 out of 139 questions (68.4%) on the 2018 PSAT/NMSQT did not require higher-order thinking skills. Essentially, 68.4% of the questions on the 2018 PSAT/NMSQT required higher-order thinking according to Webb's Depth of Knowledge on the HCRM required Recall and Reproduction (DOK Level 1) or Skills and Concepts (DOK Level 2). The data revealed a range of [1,1] Remember, Recall through [3,5] Evaluate, Strategic Thinking. The range of scoring for each PSAT/NMSQT question spanned 11 categories across the HCRMs with a mean of [1.99,3], which translates to Apply on Bloom's axis of the HCRM and an average of just under Skills and Concepts on Webb's DOK axis of the

HCRM. Additionally, the median and the mode of the scoring data also resulted in [2,3], which translates to Apply, Skills and Concepts. Moreover, the coordinates [2,3] translated to Apply, Skills and Concepts was the most frequently-used scoring cell on the HCRMs throughout the PSAT/NMSQT.

Table 4. Overall Heat Map visualizing the density of the distribution in each cell of the HCRM for all of the 2018 PSAT/NMSQT sections (Hess, 2009).

HESS COGNITIVE RIGOR MATRIX | 2018 PSAT Overall Heat Map					Scale
Bloom's Revised Taxonomy	Webb's DOK Level 1 = [1, _] Recall and Reproduction	Webb's DOK Level 2 = [2, _] Skills and Concepts	Webb's DOK Level 3 = [3, _] Strategic Thinking/Reasoning	Webb's DOK Level 4 = [4, _] Extended Thinking	
Remember = [_ , 1] Retrieve knowledge from long-term memory, recognize, recall, locate, identify	Remember, Recall [1, 1]	Use these Hess CRM curricular examples with most assignments or assessments.			1-5%
Understand = [_ , 2] Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion, predict, compare or contrast, match like ideas, explain, construct models	Understand, Recall [1,2]	Understand, Skills and Concepts [2,2]	Understand, Strategic Thinking/Reasoning [3,2]	Understand, Extended Thinking [4,2]	6-15%
Apply = [_ , 3] Carry out or use a procedure in a given situation; carry out (apply to a familiar task) or use (apply) to an unfamiliar task	Apply, Recall [1,3]	Apply, Skills and Concepts [2,3]	Apply. Strategic Thinking/Reasoning [3,3]	Apply, Extending Thinking [4,3]	16-25%
Analyze = [_ , 4] Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, fi nd coherence, deconstruct	Analyze, Recall [1,4]	Analyze, Skills and Concepts [2,4]	Analyze, Strategic Thinking/Reasoning [3,4]	Evaluate, Extended Thinking [4,4]	26-35%
Evaluate = [_ , 5] Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique	"UG" –unsubstantiated generalizations = stating an opinion without providing any support for it!		Evaulate, Strategic Thinking/Reasoning [3,5]	Evaluate, Extended Thinking [4,5]	36-45%
Create = [_ , 6] Reorganize elements into new patterns structures, generate, hypothesize, design, plan, produce	Create, Recall [1,6]	Create, Skills and Concepts [2,6]	Create, Strategic Thinking/Reasoning [3,6]	Create, Extended Thinking [4,6]	46% +

Table 5. Bar graph depicting frequency of each question scored using HCRM for all of the 2018 PSAT/NMSQT sections.

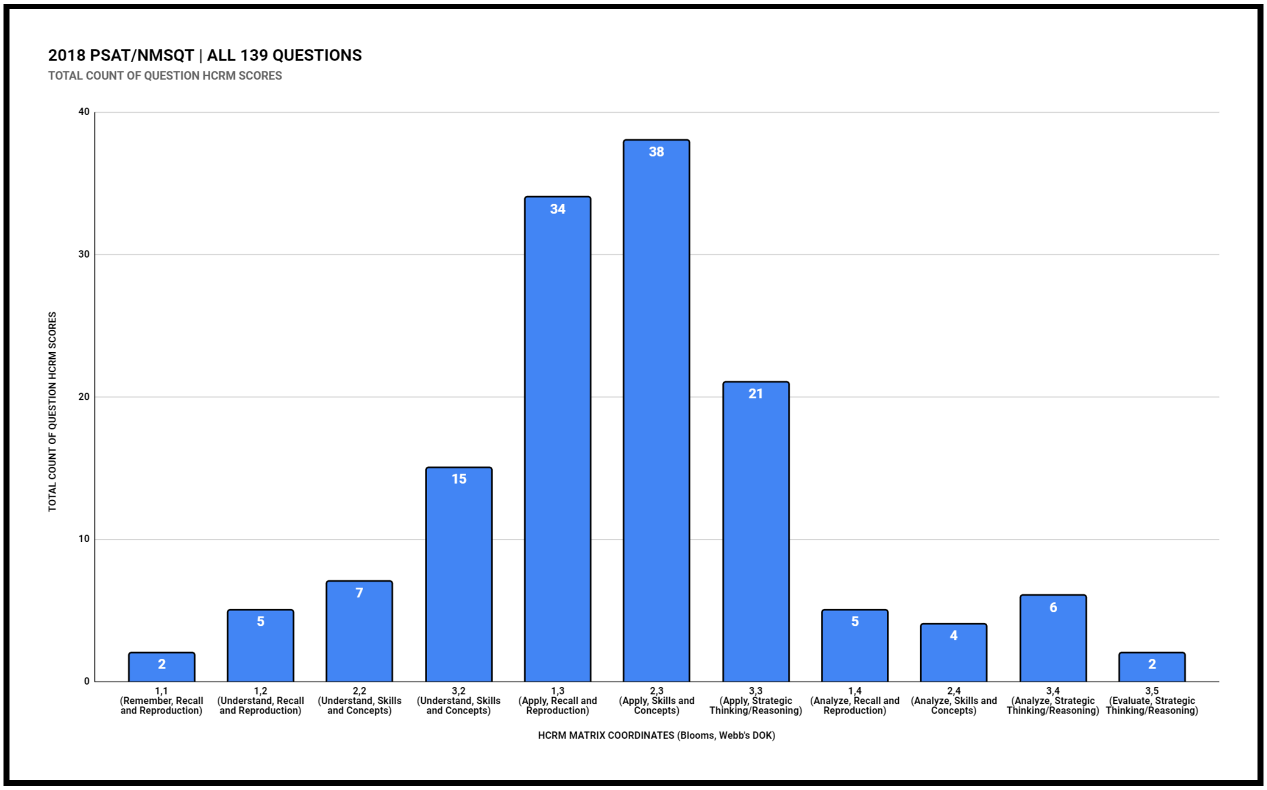
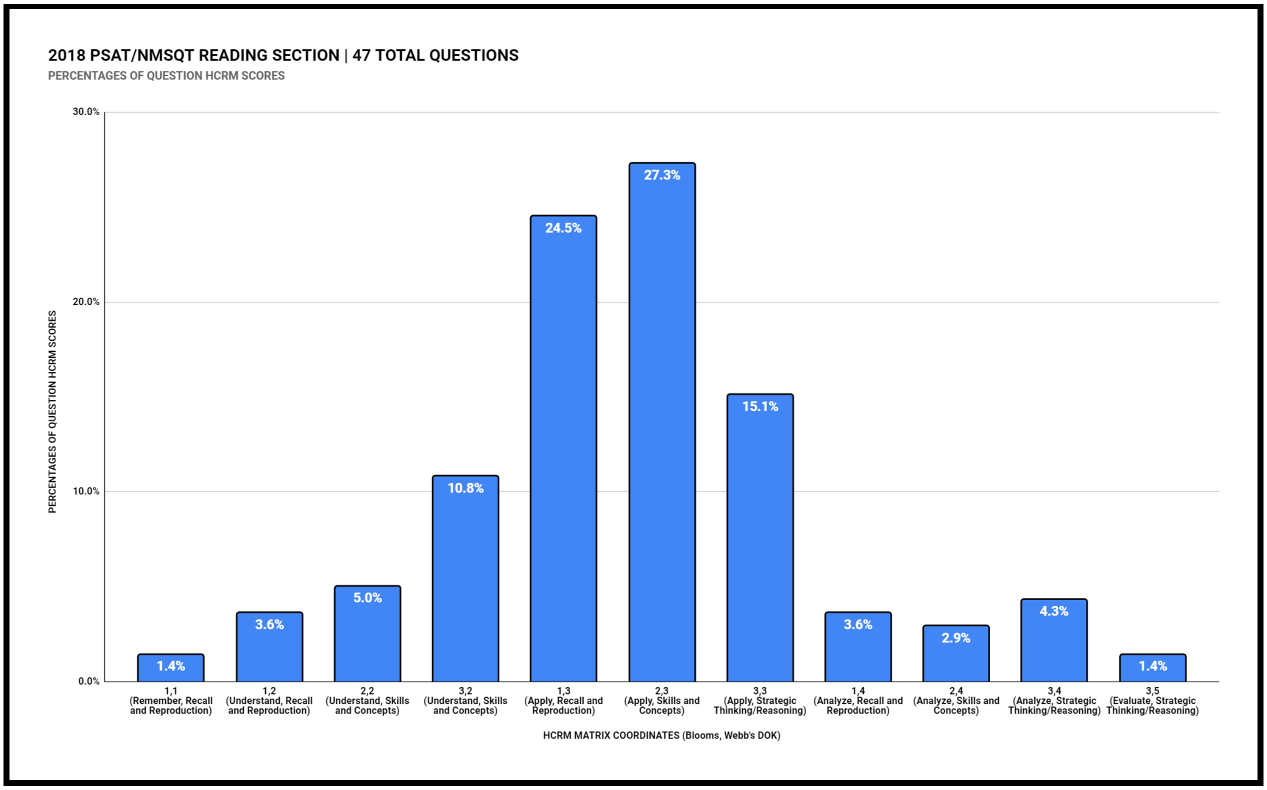


Table 6. Bar graph depicting percentage of each question scored using HCRM for all of the 2018 PSAT/NMSQT sections.



Findings for Research Question 1

1. *In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Reading Section questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?*

Precisely 27 out of the 47 questions (57.4%) on the 2018 PSAT/NMSQT Reading section contained language, that compared to the Close Reading and Listening HCRM definition of higher-order thinking. Essentially, the majority of questions in the Reading section required

Strategic Thinking and Reasoning (DOK Level 3) or Extended Thinking (DOK Level 4). Conversely, 20 out of 48 questions (42.6%) on the 2018 PSAT/NMSQT Reading section did not require higher-order thinking skills. Invariably, 42.6% of the questions on the 2018 PSAT/NMSQT Reading section required Recall and Reproduction (DOK Level 1) or Skills and Concepts (DOK Level 2). The data revealed a range of [1,1] Recall, Remember through [3,5] Strategic Thinking, Evaluate. The range of scoring for each PSAT/NMSQT question spanned ten categories across the HCRMs with a mean of [2.45,3] Apply, on Bloom's axis of the HCRM and the latter coordinate scoring solidly within the Skills and Concepts plane on Webb's DOK axis of the HCRM. Additionally, the median of the scoring data in the Reading section resulted in [2,3] Skills and Concepts, Apply. Noticeably, the mode of the Reading section data was the only section of the PSAT to have the mode fall within the higher-order thinking region of the HCRM [3,2] Strategic Thinking, Understand.

The specificity of the language used in each question on the 2018 PSAT/NMSQT aligned with the Reading section of the 2018 PSAT/NMSQT to the Close Reading and Listening HCRM quite accurately. Significantly, the traits or qualifications for each of the cells on the Close Reading and Listening HCRM closely linked to the requirements of each of the questions on the Reading section.

- Coordinate [1,1] Recall, Remember required test-takers to recall, recognize, or locate basic facts, terms, details, events or ideas explicit in texts in order to answer the questions successfully.

- Coordinate [1,2] Recall, Understand required test-takers to select appropriate words when intended meaning or definition is clearly evident in order to answer the questions successfully.
- Coordinate [2, 2] Skills and Concepts, Understand required test-takers to specify, explain, show relationships; explain why (e.g., cause-effect), summarize results, concepts ideas, or make basic inferences or logical predictions from data or texts.
- Coordinate [2,3] Skills and Concepts, Apply required test-takers to obtain and interpret information using text features to answer the question successfully.
- Coordinate [1,4] Recall, Analyze required test-takers to identify whether specific information is contained in graphic representations (e.g., map, chart, table, graph, T-chart, diagram) or text features (e.g., headings, subheadings, captions).
- Coordinate [2,4] Skills and Concepts, Analyze required test-takers to analyze format, organization, and internal text structure (signal words, transitions, semantic cues) of different texts to answer the question successfully.

Reaching the qualification of higher-order thinking skills by attending to the DOK Level 3 region of the Close Reading and Listening HCRM, Coordinate [3,2] Strategic Thinking, Understand, required test-takers to describe how word choice, point of view, or bias may affect the readers' interpretations of a text. Also, Coordinate [3,3] Strategic Thinking, Apply required test-takers to apply word choice, point of view, style to impact readers' or viewers' interpretation of a text. Furthermore, Coordinate [3,4] Strategic Thinking, Analyze required test-takers to analyze interrelationships among concepts, issues, or problems or use reasoning, planning, and evidence to support information to answer the question successfully. Lastly, Coordinate [3,5]

Strategic Thinking, Evaluate required test-takers to cite evidence and develop a logical argument for conjectures. None of the questions contained within the Reading Section reached DOK Level 4. Overall, the Reading Section had the widest range, the highest quantity and the highest percentage of questions that scored in the higher-order thinking region of the HCRM. The language found on the grade 10-11 PSAT/NMSQT Reading Section questions compares positively to Hess' Cognitive Rigor Matrix definition of higher-order thinking with the majority of questions commanding the test-taker to employ higher-order thinking skills.

Table 7. Overall Heat Map visualizing the density of the distribution in each cell of the HCRM for the 2018 PSAT/NMSQT Reading section (Hess, 2009).

HESS COGNITIVE RIGOR MATRIX (READING & LISTENING CRM) | 2018 PSAT Reading Section Heat Map					
Bloom's Revised Taxonomy	Webb's DOK Level 1 = [1, _] Recall and Reproduction	Webb's DOK Level 2 = [2, _] Skills and Concepts	Webb's DOK Level 3 = [3, _] Strategic Thinking/Reasoning	Webb's DOK Level 4 = [4, _] Extended Thinking	Scale
Remember = [_ , 1] Retrieve knowledge from long-term memory, recognize, recall, locate, identify	- Recall, recognize, or locate basic facts, terms, details, events, or ideas explicit in texts - Read words orally in connected text with fluency and accuracy **[1,1] = 2/47 questions, 4.3%**	Use these Hess CRM curricular examples with most close reading or listening assignments or assessments in any content area.			1-5%
Understand = [_ , 2] Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion, predict, compare or contrast, match like ideas, explain, construct models	- Identify or describe literary elements (characters, setting, sequence, etc.) - Select appropriate words when intended meaning or definition is clearly evident - Describe or explain who, what, where, when, or how - Define or describe facts, details, terms, principles - Write simple sentences **[1,2] = 1/47 questions, 2.1%**	- Specify, explain, show relationships; explain why (e.g., cause–effect) - Give non examples or examples - Summarize results, concepts, ideas - Make basic inferences or logical predictions from data or texts - Identify main ideas or accurate generalizations of texts - Locate information to support explicit-implicit central ideas **[2,2] = 4/47 questions, 8.5%**	- Explain, generalize, or connect ideas using supporting evidence (quote, example, text reference) - Identify or make inferences about explicit or implicit themes - Describe how word choice, point of view, or bias may affect the readers' interpretation of a text - Write multi paragraph composition for specific purpose, focus, voice, tone, and audience **[3,2] = 14/47 questions, 29.8%**	- Explain how concepts or ideas specifically relate to other content domains (e.g., social, political, historical) or concepts - Develop generalizations of the results obtained or strategies used and apply them to new problem-based situations [4,2]	6-15%
Apply = [_ , 3] Carry out or use a procedure in a given situation; carry out (apply to a familiar task) or use (apply) to an unfamiliar task	- Use language structure (pre-, or suffix) or word relationships (synonym or antonym) to determine meaning of words - Apply rules or resources to edit spelling, grammar, punctuation, conventions, word use - Apply basic formats for documenting sources [1,3]	- Use context to identify the meaning of words or phrases - Obtain and interpret information using text features - Develop a text that may be limited to one paragraph - Apply simple organizational structures (paragraph, sentence types) in writing **[2,3] = 9/47 questions, 19.1%**	- Apply a concept in a new context - Revise final draft for meaning or progression of ideas - Apply internal consistency of text organization and structure to composing a full composition - Apply word choice, point of view, style to impact readers' or viewers' interpretation of a text **[3,3] = 7/47 questions, 14.9%**	- Illustrate how multiple themes (historical, geographic, social, artistic, literary) may be interrelated - Select or devise an approach among many alternatives to research a novel problem [4,3]	16-25%
Analyze = [_ , 4] Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, fi nd coherence, deconstruct	- Identify whether specific information is contained in graphic representations (e.g., map, chart, table, graph, T-chart, diagram) or text features (e.g., headings, subheadings, captions) - Decide which text structure is appropriate to audience and purpose **[1,4] = 3/47 questions, 6.4%**	- Categorize or compare literary elements, terms, facts or details, events - Identify use of literary devices - Analyze format, organization, and internal text structure (signal words, transitions, semantic cues) of different texts - Distinguish: relevant–irrelevant information; fact or opinion - Identify characteristic text features; distinguish between texts, genres **[2,4] = 1/47 questions, 2.1%**	- Analyze information within data sets or texts - Analyze interrelationships among concepts, issues, problems - Analyze or interpret author's craft (literary devices, viewpoint, or potential bias) to create or critique a text - Use reasoning, planning, and evidence tosupport inferences **[3,4] = 5/47 questions, 10.6%**	- Analyze multiple sources of evidence, or multiple works by the same author, or across genres, time periods, themes - Analyze complex or abstract themes, perspectives, concepts - Gather, analyze, and organize multiple information sources - Analyze discourse styles [4,4]	26-35%
Evaluate = [_ , 5] Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique	"UG" –unsubstantiated generalizations = stating an opinion without providing any support for it!		- Cite evidence and develop a logical argument for conjectures - Describe, compare, and contrast solution methods - Verify reasonableness of results - Justify or critique conclusions drawn **[3,5] = 1/47 questions, 2.1%**	- Evaluate relevancy, accuracy, and completeness of information from multiple sources - Apply understanding in a novel way, provide argument or justification for the application [4,5]	36-45%
Create = [_ , 6] Reorganize elements into new patterns structures, generate, hypothesize, design, plan, produce	- Brainstorm ideas, concepts, problems, or perspectives related to a topic, principle, or concept [1,6]	- Generate conjectures or hypotheses based on observations or prior knowledge and experience [2,6]	- Synthesize information within one source or text - Develop a complex model for a given situation - Develop an alternative solution [3,6]	- Synthesize information across multiple sources or texts - Articulate a new voice, alternate them [4,6]	46% +

Table 8. Bar graph depicting frequency of each question scored using HCRM for the 2018 PSAT/NMSQT Reading section.

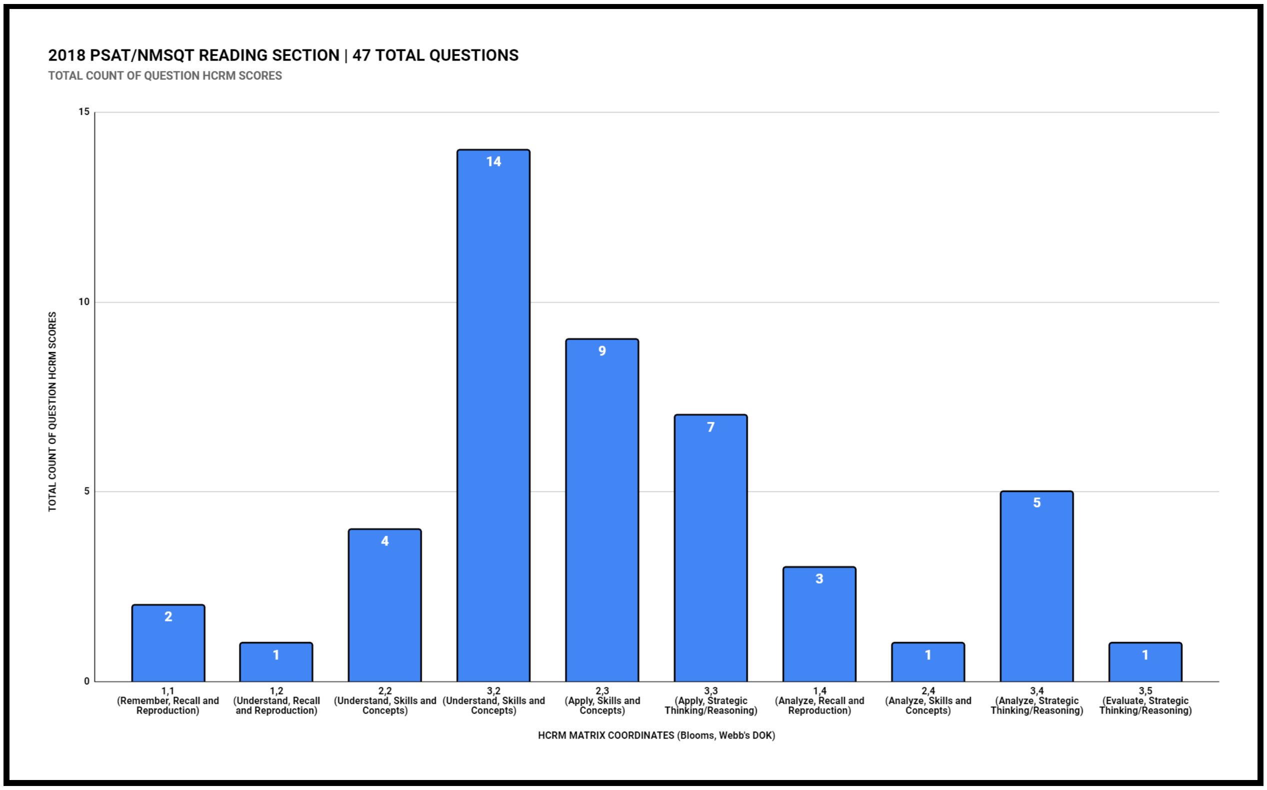
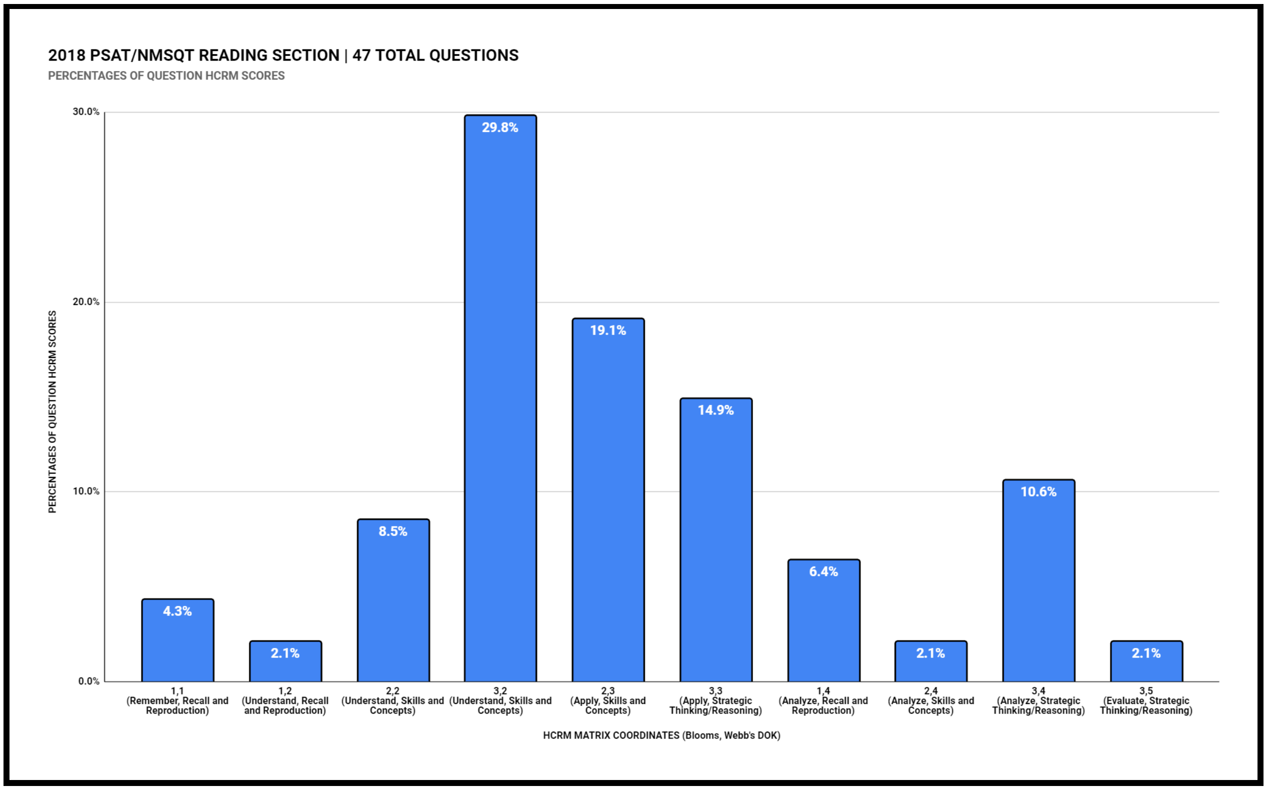


Table 9. Bar graph depicting percentage of each question scored using HCRM for the 2018 PSAT/NMSQT Reading section.



Findings for Research Question 2

2. *In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Math section (both calculator and non-calculator) questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?*

Comprehensively, four out of the 48 questions (8.3%) on the 2018 PSAT/NMSQT Math Sections (calculator and non-calculator) contained language as compared to the Math and Science HCRM definition of higher-order thinking. Essentially, 8.3% of the questions required

Strategic Thinking and Reasoning (DOK Level 3) or Extended Thinking (DOK Level 4), according to Webb's Depth of Knowledge on the Math and Science HCRM. Thereupon, 44 out of 48 questions (91.7%) on the 2018 PSAT/NMSQT Math sections (calculator and non-calculator) did not require higher-order thinking skills. In essence, 91.7% of the questions on the 2018 PSAT/NMSQT Math sections (calculator and non-calculator) required to recall and Reproduction (DOK Level 1) or skills and concepts (DOK Level 2), according to Webb's Depth of Knowledge on the Math and Science HCRM. The data revealed a range of [1,2] Recall, Understand through [3,3] Strategic Thinking, Apply. The scoring range for each PSAT/NMSQT question spanned eight categories across the HCRMs with a mean of [1.89,3]. Additionally, the median and the mode of the scoring data also resulted in [2,3] Skills and Concepts, Apply. Similar to the overall mapping of the scoring, the coordinates within the Math Sections fell into [2,3] Skills and Concepts, Apply by 45.2 percentage points as compared to the next highest HCRM scoring category.

The specificity of the language used in each question on the 2018 PSAT/NMSQT aligned to the Math and Science HCRM intuitively. As the coders were measuring each question's requirements of the test-taker to complete the question successfully, the coders discovered patterns of language that fell directly within the traits detailed in each cell of the Math and Science HCRM.

- Coordinate [1,2] Recall, Understand asked test-takers to solve a one step problem or evaluate an expression.
- Coordinate [2,2] Skills and Concepts, Understand asked test-takers to specify and explain relationships or make basic inferences or logical predictions from data or observations.

- Coordinate [1,3] Recall, Apply asked test-takers to solve linear equations, apply an algorithm or a formula, or follow simple procedures.
- The most frequently scored Coordinate [2,3] Skills and Concepts, Apply asked students to solve routine problems by applying multiple concepts or decision points or retrieve information from a table, graph or figure and use it to solve a problem requiring multiple steps.
- Coordinate [1,4] Recall, Analyze asked test-takers to retrieve information from a table or graph to answer a question.
- Coordinate [2,4] Skills and Concepts, Analyze asked test-takers to interpret data from a simple graph.

Reaching the qualification of higher-order thinking skills by achieving the requirements of the DOK Level 3 region of the Math and Science HCRM, Coordinate [3,2] Strategic Thinking, Understand asked test-takers to explain, generalize, or connect ideas using supporting evidence. Coordinate [3,3] asked test-takers to use concepts to solve non-routine problems on three total questions. None of the questions contained within the Math Sections (both the calculator and non-calculator sections) reached DOK Level 4. Overall, the language found on the grade 10-11 PSAT/NMSQT Math section (both calculator and non-calculator) questions do not match Hess' Cognitive Rigor Matrix definition of higher-order thinking, with the vast majority of questions failing to require the test-taker to employ higher-order thinking skills.

Table 10. Overall Heat Map visualizing the density of the distribution in each cell of the HCRM for the 2018 PSAT/NMSQT Math section (Hess, 2009).

HESS COGNITIVE RIGOR MATRIX (MATH SCIENCE CRM) | 2018 PSAT Math, Calculator and Non-Calculator Sections Heat Map					Scale
Bloom's Revised Taxonomy	Webb's DOK Level 1 = [1, _] Recall and Reproduction	Webb's DOK Level 2 = [2, _] Skills and Concepts	Webb's DOK Level 3 = [3, _] Strategic Thinking/Reasoning	Webb's DOK Level 4 = [4, _] Extended Thinking	
Remember = [_ , 1] Retrieve knowledge from long-term memory, recognize, recall, locate, identify	- Recall, observe, recognize facts, principles, properties - Recall/identify conversions among representations or numbers (e.g., customary and metric measures) [1,1]	Use these Hess CRM curricular examples with most mathematics or science assignments or assessments.			1-5%
Understand = [_ , 2] Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion, predict, compare or contrast, match like ideas, explain, construct models	- Evaluate an expression - Locate points on a grid or number on number line - Solve a one-step problem - Represent math relationships in words, pictures, or symbols - Read, write, compare decimals in scientific notation **[1,2] = 2/48 questions, 4.2%**	- Specify and explain relationships (e.g., nonexamples or examples; cause-effect) - Make and record observations - Explain steps followed - Summarize results or concepts - Make basic inferences or logical predictions from data or observations - Use models diagrams to represent or explain mathematical concepts - Make and explain estimates **[2,2] = 3/48 questions, 6.3%**	- Use concepts to solve nonroutine problems - Explain, generalize, or connect ideas using supporting evidence - Make and justify conjectures - Explain thinking or reasoning when more than one solution or approach is possible - Explain phenomena in terms of concepts **[3,2] = 1/48 questions, 2.1%**	- Relate mathematical or scientifi c concepts to other content areas, other domains, or other concepts - Develop generalizations of the results obtained and the strategies used (from investigation or readings) and apply them to new problem situations [4,2]	6-15%
Apply = [_ , 3] Carry out or use a procedure in a given situation; carry out (apply to a familiar task) or use (apply) to an unfamiliar task	- Follow simple procedures (recipe-type directions) - Calculate, measure, apply a rule (e.g., rounding) - Apply algorithm or formula (e.g., area, perimeter) - Solve linear equations - Make conversions among representations or numbers, or within and between customary and metric measures **[1,3] = 6/48 questions, 12.5%**	**- Select a procedure according to criteria and perform it - Solve a routine problem applying multiple concepts or decision points - Retrieve information from a table, graph, or figure and use it to solve a problem requiring multiple steps - Translate between tables, graphs, words, and symbolic notations (e.g., graph data from a table) - Construct models given criteria [2,3] = 29/48 questions, 60.4%**	- Design an investigation for a specific purpose or research question - Conduct a designed investigation - Use concepts to solve nonroutine problems - Use show reasoning, planning, and evidence - Translate between problem and symbolic notation when not a direct translation **[3,3] = 3/48 questions, 6.3%**	- Select or devise an approach among many alternatives to solve a problem - Conduct a project that specifies a problem, identifies solution paths, solves the problem, and reports results [4,3]	16-25%
Analyze = [_ , 4] Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, fi nd coherence, deconstruct	- Retrieve information from a table or graph to answer a question - Identify whether specifi c information is contained in graphic representations (e.g., table, graph, T-chart, diagram) - Identify a pattern trend **[1,4] = 2/48 questions, 2.1%**	- Categorize, classify materials, data, fi gures based on characteristics - Organize or order data - Compare−contrast figures or data - Select an appropriate graph and organize - Interpret data from a simple graph - Extend a pattern **[2,4] = 2/48 questions, 2.1%**	- Compare information within or across data sets or texts - Analyze and draw conclusions from data, citing evidence - Generalize a pattern - Interpret data from complex a graph - Analyze similarities differences between procedures or solutions [3,4]	- Analyze multiple sources of evidence - Analyze complex abstract themes - Gather, analyze, and evaluate information [4,4]	26-35%
Evaluate = [_ , 5] Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique	"UG" −unsubstantiated generalizations = stating an opinion without providing any support for it!		- Cite evidence and develop a logical argument for concepts or solutions - Describe, compare, and contrast solution methods - Verify reasonableness of results [3,5]	- Gather, analyze, and evaluate information to draw conclusions - Apply understanding in a novel way, provide argument or justification for the application [4,5]	36-45%
Create = [_ , 6] Reorganize elements into new patterns structures, generate, hypothesize, design, plan, produce	- Brainstorm ideas, concepts, or perspectives related to a topic [1,6]	Generate conjectures or hypotheses based on observations or prior knowledge and experience [2,6]	- Synthesize information within one data set, source, or text - Formulate an original problem given a situation - Develop a scientific or mathematical model for a complex situation [3,6]	- Synthesize information across multiple sources or texts - Design a mathematical model to inform and solve a practical or abstract situation [4,6]	46% +

Table 11. Bar graph depicting frequency of each question scored using HCRM for the 2018 PSAT/NMSQT Math section.

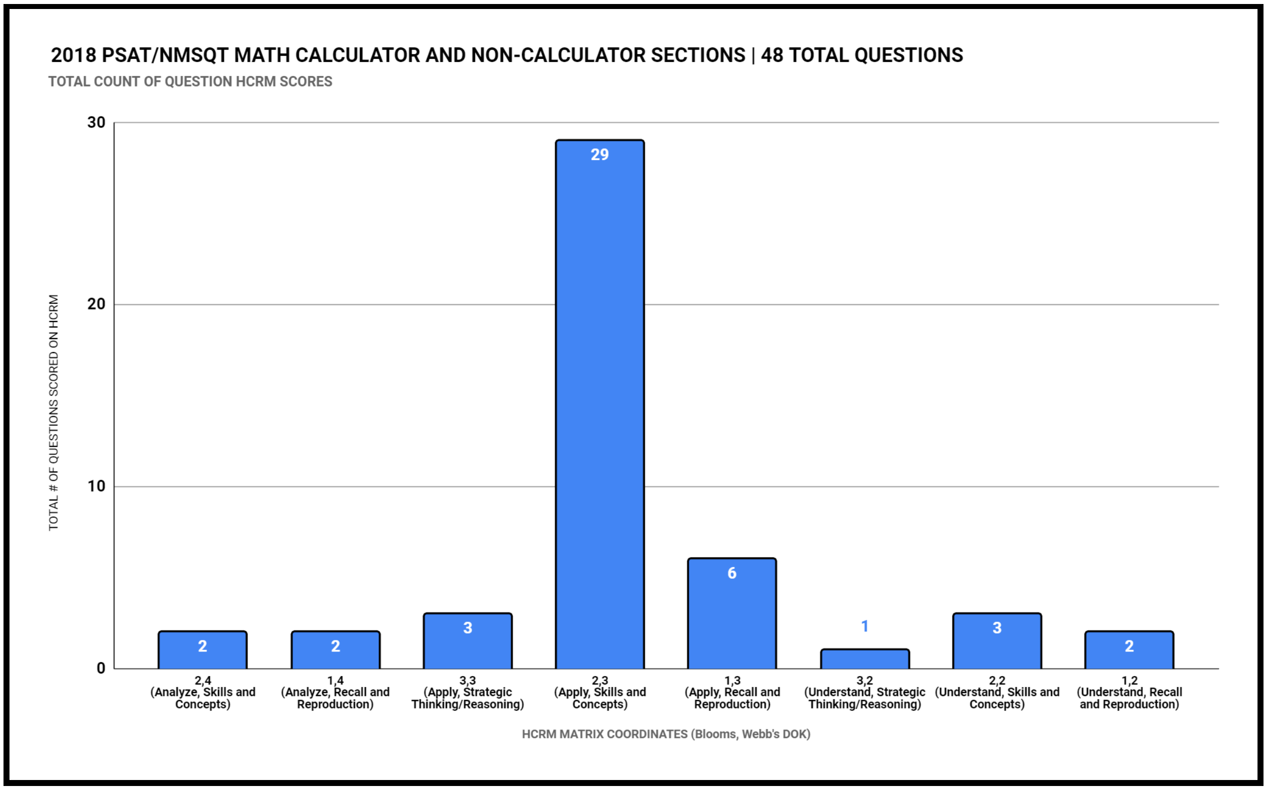
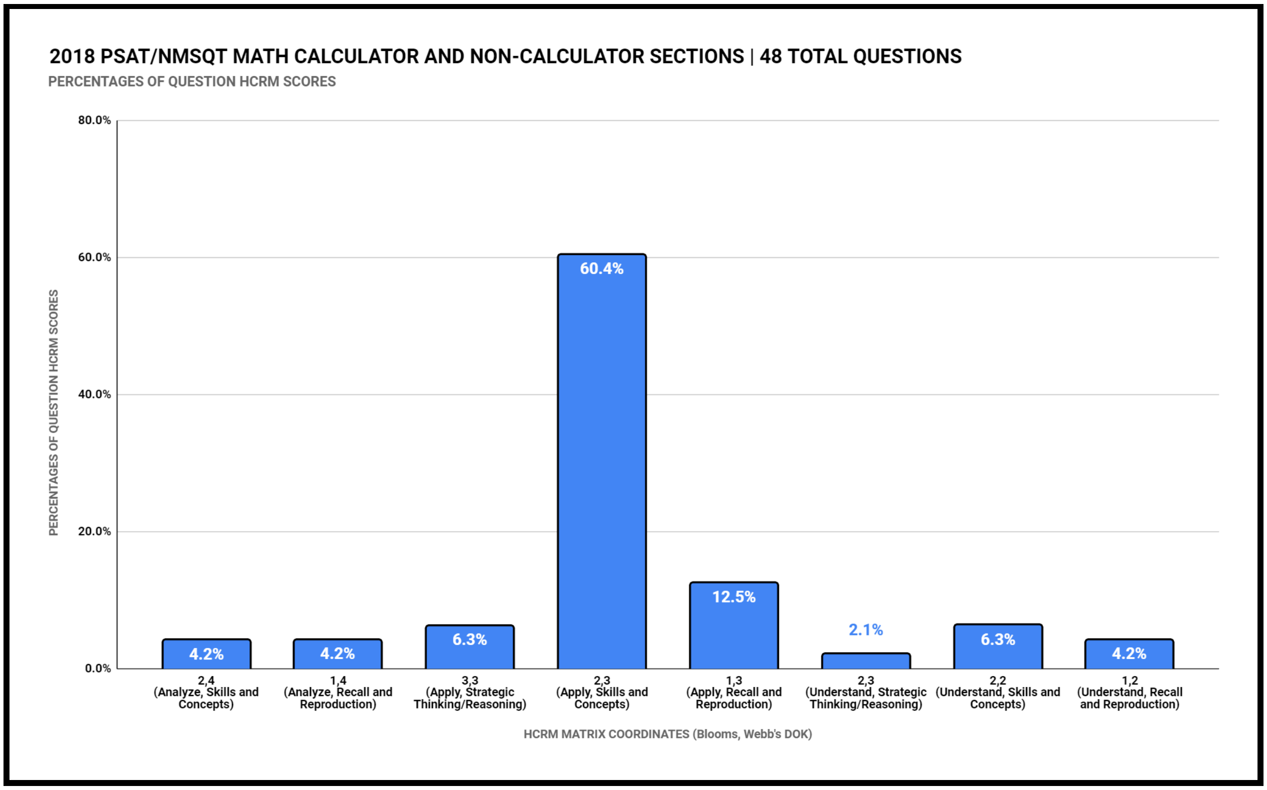


Table 12. Bar graph depicting percentage of each question scored using HCRM for the 2018 PSAT/NMSQT Math section.



Findings for Research Question 3

3. *In what way(s) does the language found on the grade 10-11 PSAT/NMSQT Writing section's questions compare to Hess' Cognitive Rigor Matrix definition of higher-order thinking?*

A total of 13 out of the 44 questions (29.5%) on the 2018 PSAT/NMSQT Writing section contained language compared to the Written and Oral Communication HCRM definition of higher-order thinking. Essentially, 29.5% of the questions required Strategic Thinking and Reasoning (DOK Level 3) or Extended Thinking (DOK Level 4), according to Webb's Depth of Knowledge of the Close Reading and Listening HCRM. Conversely, 31 out of 44 questions

(70.5%) on the 2018 PSAT/NMSQT Writing section did not require higher-order thinking skills. Invariably, 70.5% of the questions on the 2018 PSAT/NMSQT Writing section required Recall and Reproduction (DOK Level 1) or skills and concepts (DOK Level 2). The data revealed a range of [1,1] Remember, Recall through [3,4] Analyze, Strategic Thinking. The range of scoring for each PSAT/NMSQT Writing section question spanned five categories across the HCRMs with a mean of [1.61,3] Apply, on Bloom's axis of the HCRM and the latter coordinate scoring solidly within the Skills and Concepts plane on Webb's DOK axis of the HCRM. Additionally, the median of the scoring data in the Writing section resulted in [1,3] Recall, Apply and the mode of the Writing section was also [1,3] Recall, Apply.

The nature of the language used in each question on the 2018 PSAT/NMSQT aligned from the Writing section of the 2018 PSAT/NMSQT to the Written and Oral Communication HCRM quite accurately as the researchers deciphered each of the questions. Significantly, the traits or qualifications for each of the cells on the Written and Oral Communication HCRM closely linked to the requirements of each of the questions on the Writing section.

- Coordinate [1,2] Recall, Understand required test-takers to select appropriate words when intended meaning or definition is clearly evident in order to answer the questions successfully.
- Coordinate [1,3] Recall, Apply required test-takers to apply rules or use resources to edit specific spelling, grammar, punctuation, conventions or word use to answer the prompts successfully.

- Coordinate [2,4] Skills and Concepts, Analyze required test-takers to analyze format, organization, and internal text structure (signal words, transitions, semantic cues) of different texts to answer the question successfully.

Reaching the qualification of higher-order thinking skills by achieving a coordinate score within the DOK Level 3 region of the Written and Oral Communication HCRM, Coordinate [3,3] Apply, Strategic Thinking required test-takers to describe how word choice, point of view, or bias may affect the readers' interpretations of a text. In addition, Coordinate [3,4] Analyze, Strategic Thinking also reached the DOK Level 3 region by requiring test-takers to analyze interrelationships among concepts, issues, or problems or use reasoning, planning, and evidence to support information to answer the question successfully. None of the questions contained within the Writing section reached DOK Level 4. Overall, the language found on the grade 10-11 PSAT/NMSQT Writing section's questions compares incompatibly to Hess' Cognitive Rigor Matrix definition of higher-order thinking with the majority of questions not requiring the test-taker to employ higher-order thinking skills.

Table 13. Overall Heat Map visualizing the density of the distribution in each cell of the HCRM for the 2018 PSAT/NMSQT Writing section (Hess, 2009).

HESS COGNITIVE RIGOR MATRIX (WRITTEN & ORAL COMMUNICATION CRM) | 2018 PSAT Writing Section Heat Map					Scale
Bloom's Revised Taxonomy	Webb's DOK Level 1 = [1, _] Recall and Reproduction	Webb's DOK Level 2 = [2, _] Skills and Concepts	Webb's DOK Level 3 = [3, _] Strategic Thinking/Reasoning	Webb's DOK Level 4 = [4, _] Extended Thinking	
Remember = [_ , 1] **Retrieve knowledge from long-term memory, recognize, recall, locate, identify**	- Complete short answer questions with facts, details, terms, principles, etc. (e.g., label parts of diagram) [1,1]	Use these Hess CRM curricular examples with most writing and oral communication assignments or assessments in any content area.			1-5%
Understand = [_ , 2] **Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion, predict, compare or contrast, match like ideas, explain, construct models**	- Describe or define facts, details, terms, principles, etc. - Select appropriate word or phrase to use when intended meaning or definition is clearly evident - Write simple complete sentences o Add an appropriate caption to a photo or illustration - Write "fact statements" on a topic (e.g., spiders build webs) **[1,2] = 2/44 questions, 4.5%**	- Specify, explain, show relationships; explain why, cause–effect - Provide and explain non examples and examples - Take notes; organize ideas or data (e.g., relevance, trends, perspectives) - Summarize results, key concepts, ideas - Explain central ideas or accurate generalizations of texts or topics - Describe steps in a process (e.g., science procedure, how to and why control variables) [2,2]	- Write a multi paragraph composition for specific purpose, focus, voice, tone, and audience - Develop and explain opposing perspectives or connect ideas, principles, or concepts using supporting evidence (quote, example, text reference, etc.) - Develop arguments of fact (e.g., Are these criticisms supported by the historical facts? Is this claim or equation true?) [3,2]	- Use multiple sources to elaborate on how concepts or ideas specifically draw from other content domains or differing concepts (e.g., research paper, arguments of policy—should this law be passed? What will be the impact of this change?) - Develop generalizations about the results obtained or strategies used and apply them to a new problem or contextual scenario [4,2]	6-15%
Apply = [_ , 3] **Carry out or use a procedure in a given situation; carry out (apply to a familiar task) or use (apply) to an unfamiliar task**	**- Apply rules or use resources to edit specific spelling, grammar, punctuation, conventions, or word use** **- Apply basic formats for documenting sources** **[1,3] = 28/47 questions, 63.6%**	- Use context to identify or infer the intended meaning of words or phrases - Obtain, interpret, and explain information using text features (table, diagram, etc.) - Develop a (brief) text that may be limited to one paragraph, précis - Apply basic organizational structures (paragraph, sentence types, topic sentence, introduction, etc.) in writing [2,3]	- Revise final draft for meaning, progression of ideas, or logic chain - Apply internal consistency of text organization and structure to a full composition or oral communication - Apply a concept in a new context - Apply word choice, point of view, style, rhetorical devices to impact readers' interpretation of a text **[3,3] = 11/47 questions, 25%**	- Select or devise an approach among many alternatives to research and present a novel problem or issue - Illustrate how multiple themes (historical, geographic, social) may be interrelated within a text or topic [4,3]	16-25%
Analyze = [_ , 4] **Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, fi nd coherence, deconstruct**	- Decide which text structure is appropriate to audience and purpose (e.g., compare–contrast, proposition–support) - Determine appropriate, relevant key words for conducting an Internet search or researching a topic [1,4]	- Compare–contrast perspectives, events, characters, etc. - Analyze–revise format, organization, and internal text structure (signal words, transitions, semantic cues) of different print and non print texts - Distinguish: relevant–irrelevant information; fact–opinion (e.g., What are the characteristics of a hero's journey?) - Locate evidence that supports a perspective–differing perspectives **[2,4] = 1/47 questions, 2.3%**	- Analyze interrelationships among concepts, issues, and problems in a text - Analyze impact or use of author's craft (literary devices, viewpoint, dialogue) in a single text - Use reasoning and evidence to generate criteria for making and supporting an argument of judgment (Was FDR a great president? Who was the greatest ball player?) - Support conclusions with evidence **[3,4] = 2/47 questions, 4.5%**	- Analyze multiple sources of evidence, or multiple works by the same author, or across genres, or time periods - Analyze complex or abstract themes, perspectives, concepts - Gather, analyze, and organize multiple information sources - Compare and contrast conflicting judgments or policies (e.g., Supreme Court decisions) [4,4]	26-35%
Evaluate = [_ , 5] **Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique**	"UG" –unsubstantiated generalizations = stating an opinion without providing any support for it!		- Evaluate validity and relevance of evidence used to develop an argument or support a perspective - Describe, compare, and contrast solution methods - Verify or critique the accuracy, logic, and reasonableness of stated conclusions or assumptions [3,5]	- Evaluate relevancy, accuracy, and completeness of information across multiple sources - Apply understanding in a novel way, provide argument or justification for the application - Critique the historical impact (policy, writings, discoveries, etc.) [4,5]	36-45%
Create = [_ , 6] **Reorganize elements into new patterns structures, generate, hypothesize, design, plan, produce**	- Brainstorm facts, ideas, concepts, problems, or perspectives related to a topic, text, idea, issue, or concept [1,6]	- Generate conjectures, hypotheses , or predictions based on facts, observations, evidence/observations, or prior knowledge and experience - Generate believable "grounds" (reasons) for an opinion-argument [2,6]	- Develop a complex model for a given situation or problem - Develop an alternative solution or perspective to one proposed (e.g., debate) [3,6]	- Synthesize information across multiple sources or texts in order to articulate a new voice, alternate theme, new knowledge or nuanced perspective [4,6]	46% +

Table 14. Bar graph depicting frequency of each question scored using HCRM for the 2018 PSAT/NMSQT Writing section.

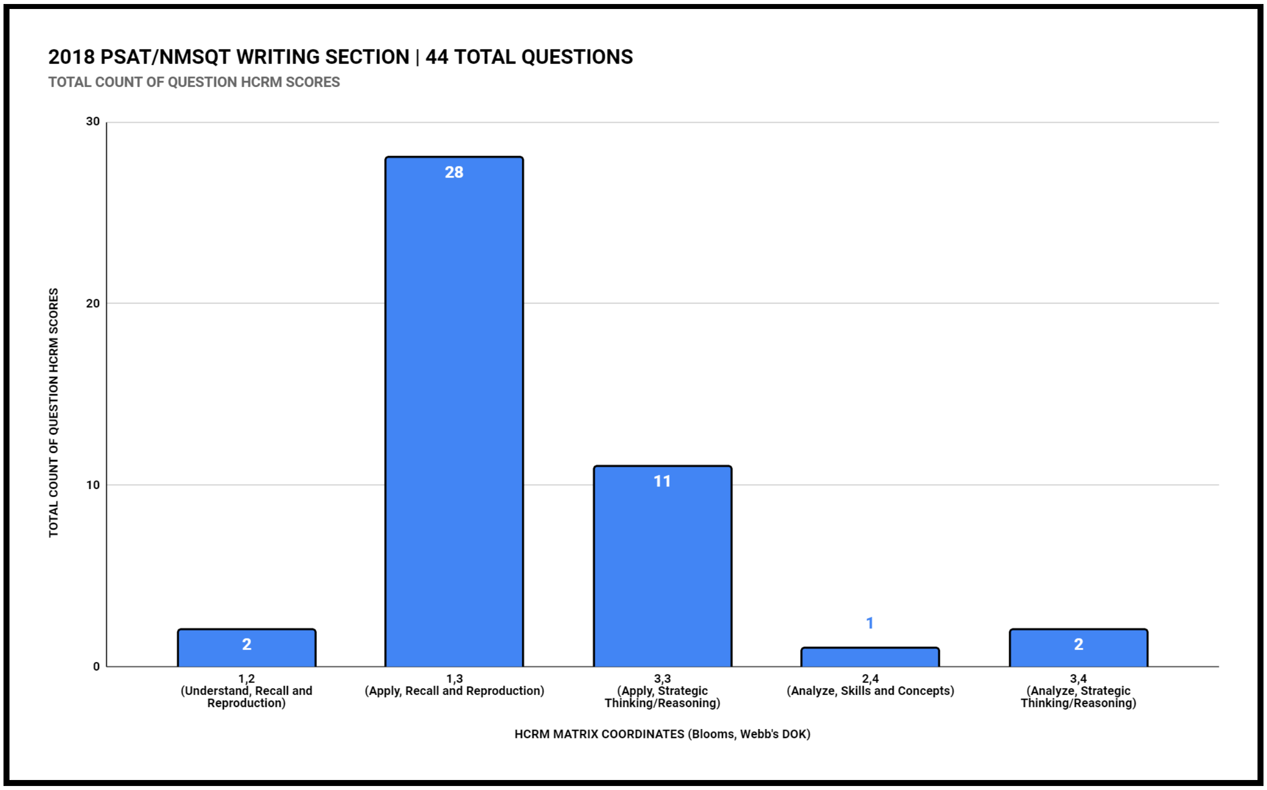
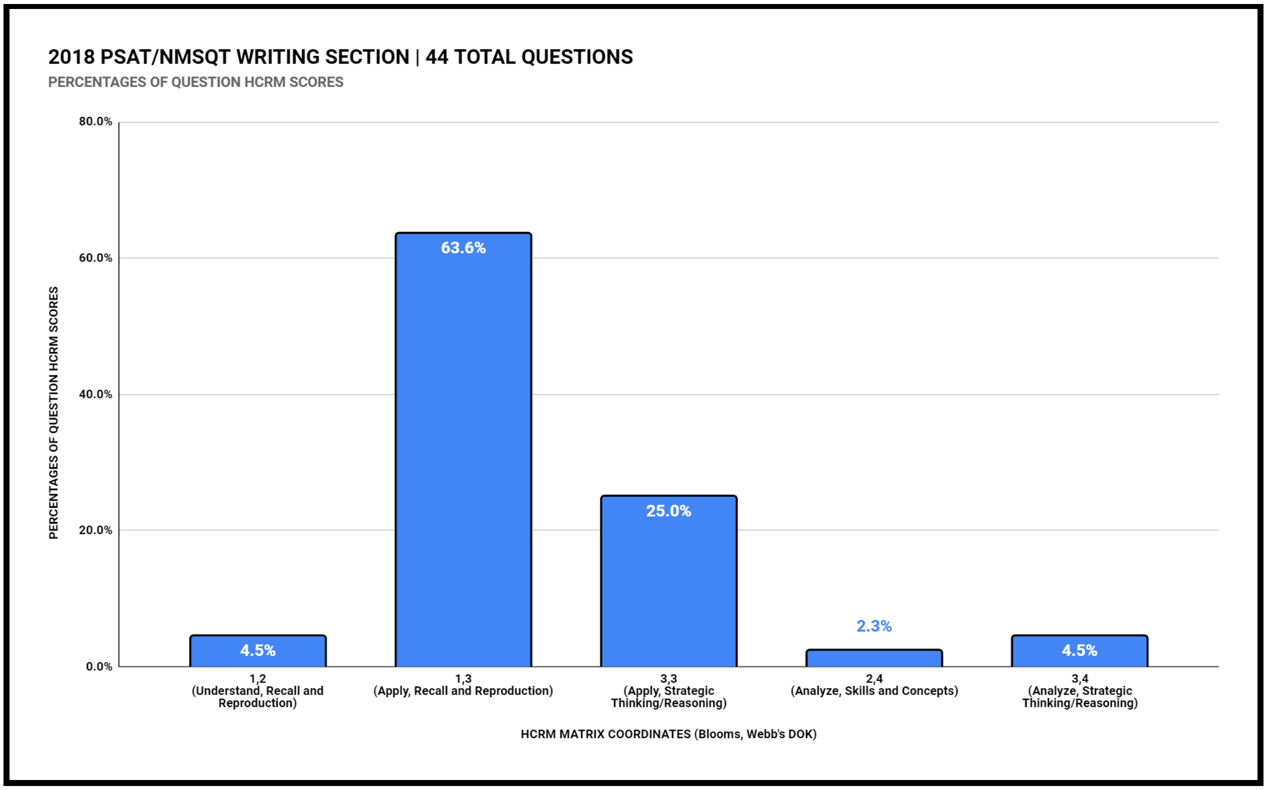


Table 15. Bar graph depicting percentage of each question scored using HCRM for the 2018 PSAT/NMSQT Writing section.



Conclusion

This study's focus was to parse out, analyze, grapple with, and compare the language of every single question on the 2018 PSAT/NMSQT to the concordant, subject-based Hess' Cognitive Rigor Matrix. As the questions were analyzed according to what was required (in terms of skills utilized and cognitive rigor accessed) for the test-taker to arrive at the answer of the question successfully. The language contained within each of the questions as defined by the subject-specific HCRM that scored a Level 3 or Level 4 designation on Webb's DOK axis earned higher-order thinking classification. The majority (57.4%) of the questions within the reading section earned higher-order thinking classification, but the majority (91.7%) of the questions on the Math section and the majority (70.5%) of the questions within the Writing Section did not earn higher-order thinking classification. Of equal importance, Coordinate [2,3] Skills and Concepts, Apply was the most frequently scored classification of questions throughout the 2018 PSAT/NMSQT. In fact, just two of the Coordinates [2,3] and [1,3] accounted for 51.8% of the questions on the 2018 PSAT/NMSQT. Furthermore, the central tendency tabulations of the data scoring resulted in Coordinate [2,3] Skills and Concepts, Apply being the same for the mean, median, and mode. In summary, none of the questions achieved Webb's DOK Level 4 and none of the questions achieved Revised Bloom's Taxonomy Level 6; Create status Extended Thinking status. Ultimately, only 31.7% of the questions scored reached the level of higher-order thinking status on the respective Hess' Cognitive Rigor Matrices.

Chapter V has been constructed to provide an overview of the research and illustrates the data of the study through the perspective of the research questions. Policy, practice, and future research are also contained within Chapter V to provide the closing details of the study.

Chapter V

CONCLUSION

Introduction

Chapter V is designed to review the overall findings of the research, referencing the data of the study as it relates to the research questions, possible outcomes for policy and practice at the local, state, and perhaps the national level of public education, and how this study can be used as a springboard for future research. The qualitative document analysis apparatus used by the researchers focused on the precise language of the 2018 PSAT/NMSQT compared to the subject-specific (math, reading, and writing) Hess' Cognitive Rigor Matrices (HCRMs). The researchers were devoted to quantifying exactly which higher-order thinking skills and levels of cognitive rigor were required of a test-taker to complete each question successfully. Accordingly, the language of every question was intensely scrutinized as compared to the traits detailed in each cell of the HCRM. Hess has created a subject-specific HCRM that is directly aligned to each section (math, reading, and writing) of the 2018 PSAT/NMSQT. As a result, the researchers effectively and accurately evaluated and scored each question to the concordant HCRM. Surprisingly, this is the first study in which the language of each PSAT/NMSQT question was paced through a HCRM.

Methodology Summary

The cornerstones in which the evaluation of the researchers laid the foundation of the study were the subject-specific Hess' Cognitive Rigor Matrices that transpose Revised Bloom's taxonomy with Webb's Depth of Knowledge. The researchers intentionally chose Hess'

Cognitive Rigor Matrix (HCRM) with the ultimate goal of taking full advantage of the HCRMs ability to be applied to different subjects without losing integrity or coherence in scoring questions. With the 2018 PSAT/NMSQT containing math, reading, and writing section, the researchers needed a measurement tool that could remain valid regardless of the subject area. Effectively, the researchers applied the content-specific descriptors in each HCRMs to categorize and plan for various levels of detail regarding the mental processing required of the test-taker working to solve the questions and learning tasks (Hess, 2014).

Webb's Depth of Knowledge, as acclimatized to the HCRMs, categorizes tasks according to the complexity of thinking required to successfully complete them (Aungst, 2014). Webb's Depth of Knowledge (DOK) Level 1; Recall and Reproduction: Tasks at this level require recall of facts or rote application of simple procedures. The task does not require any cognitive effort beyond remembering the proper response or formula (Aungst, 2014). Within the 2018 PSAT/NMSQT, test-takers were asked to remember grammar rules, reproduce information from a graph, or select appropriate word or phrase to use when the intended meaning or definition is clearly evident. Questions reaching the realm of DOK Level 2; Skills and Concepts required test-takers to make some decisions about his or her approach. Tasks with more than one mental step, such as comparing, organizing, summarizing, predicting, and estimating, are usually DOK Level 2 (Aungst, 2014). For example, the 2018 PSAT/NMSQT asked test-takers to summarize results or key concepts, obtain, interpret, and explain information using text features, or solve routine problems by applying multiple concepts or decision points. The higher-order thinking region of the HCRM includes questions that asked DOK Level 3; Strategic Thinking type of questions, in which the test-taker must use planning and evidence and thinking is more abstract. A task with multiple valid responses, where students must justify their choices, would be DOK Level 3 (Aungst, 2014). For instance, the 2018 PSAT/NMSQT delivered questions that required test-takers to use concepts to solve non-routine problems, apply word choice, point of view, style, and theoretical devices to impact reader's interpretation of a text, or describe, cite evidence, and develop a logical argument for conjectures. Interestingly, the highest level of the HCRM is the DOK Level 4. DOK Level 4 corresponds to Extended Thinking, for which test-takers are asked to perform the most complex cognitive effort. In essence, students are compelled to synthesize information from multiple sources, often over an extended period of time, or transfer knowledge from one domain to solve problems in another (Aungst, 2014) at the highest level of Webb's DOK. The 2018 PSAT/NMSQT did not have any questions that attained DOK Level 4 status. Of note, Bloom's Taxonomy is the y-axis portion of the HCRM and is applied to assessments as a classification system that defines and distinguishes different levels of human cognition—i.e., thinking, learning, and understanding. Educators have typically used Bloom's taxonomy to inform or guide the development of assessments (tests and other evaluations of student learning), curriculum (units, lessons, projects, and other learning activities), and instructional methods such as questioning strategies (The Glossary of Education Reform, 2014).

The levels of Bloom's Taxonomy embedded into HCRM include, from lowest to highest, remember, understand, apply, analyze, evaluate, and create. These levels increase in complexity based on the tasks that students are being asked to complete. The complexity of tasks does not correspond to cognitive complexity in that tasks with lower cognitive complexity in Webb's DOK could be placed into a higher level of complexity of the task in Bloom's Taxonomy.

(Sydoruk, 2018). For example, a test-taker in the math section may be asked to retrieve information from a table or graph to answer a question. Subsequently, the test-taker is asked to analyze the graph (Bloom's Level 4) but to only retrieve information (Webb's DOK Level 1). By superimposing both models, questions on the 2018 PSAT/NMSQT have the opportunity to be scored through the lens of the HCRM with the foundation of depth of understanding and the skills required to complete the task at its core.

Functionally, the qualitative document analysis apparatus allowed the researchers to practically apply Hess' Cognitive Rigor Matrices' definition of higher-order thinking skills to the language found at the core of each question on the PSAT/NMSQT. Invariably, the researchers aligned their evaluation process to the standards and expectations of a deductive category application that is outlined by Mayring (2000). Furthermore, each of the questions was evaluated through the concordant, subject-based Hess' Cognitive Rigor Matrix lens. The outcome of each evaluation of each question was a score or a coordinate. For example, the crosswalk of a question that required the test-taker to have Level 3 Webb's DOK (Strategic Thinking) and Bloom's Taxonomy of Level 2 (Understanding) within the PSAT Reading section resulted in a coordinate of [2,3] that falls within the Higher-Order Thinking region of the HCRM. To that end, interrater reliability was a priority through the evaluation portion of the study. The researchers performed the rating of each question in a double-rater, read-behind model that also gave the researchers the ability to discuss the scoring and rationale behind each of the scores. Over the course of a six-week time period, the researchers met in person to examine the 2018 PSAT/NMSQT. These methods for buttressing reliability of the examination protocol of the 2018 PSAT/NMSQT were inspired by guideposts set forth by dissertations on similar topics, embraced Webb's Alignment Tool for calibration, and gained insight from Hess's *A Local Assessment Toolkit to Promote Deeper Learning: Transforming Research Into Practice* (2018) and *Deeper Competency-Based Learning: Making Equitable, Student-Centered, Sustainable Shifts* (2020). One of the challenges encountered by the coders was that a few of the categories on the subject-specific matrices had terminology within adjacent cell coordinates that either directly or very nearly overlapped from one sector to another. For example, within the Math-Science CRM cell Coordinate [1,4] details a process in which a student retrieves information from a table or graph to answer a question, and cell Coordinate [2,3] details a process in which a student retrieves information from a table, graph, or figure and use it to solve a problem. One can see that the two traits within each distinct cell require nearly the same process and pose a challenge in accurately scoring a prompt when evaluating a prompt on a high-stakes standardized assessment. As a result, the scoring process, established long before the coding even began, entailed the coders scoring a prompt on the 2018 PSAT/NMSQT with the higher-order thinking cell coordinate when a mutual agreement could not be established or if the language blended across two cells on the HCRM. Lastly, the HCRM for the Math and Reading sections had some questions that fell in between or across coordinates. Using all of the possible and effective tools available to the researchers, experts in the field of Math and English Language arts were consulted to assist in the deductive scoring process to maximize the accuracy of the scoring of the 2018 PSAT/NMSQT.

Summary and Discussion Findings

The parent company of the 2018 PSAT/NMSQT is College Board. On the College Board website, the PSAT/NMSQT is marketed as a tool that illustrates a high school student's readiness for college (College Board, 2021). Research has shown that higher scores on the PSAT and the SAT correlate to higher first-year GPAs at the college level of academics, higher first-year retention rates, and higher graduation rates (University of California Academic Senate, 2020). In a parallel process, the PSAT/NMSQT has been found to provide students with the information necessary to make adjustments in Grade 9, 10, and 11 to help better prepare for the SAT and for the rigor of college courses (Proctor, Wyatt, & Riley, 2010). Moreover, the New Jersey Department of Education is an ardent proponent of New Jersey High School graduates showing proficiency in college and career readiness skills that include:

> Applying appropriate academic and technical skills, communicate clearly and effectively and with reason. consider the environmental, social and economic impacts of decisions, demonstrate creativity and innovation, employ valid and reliable research strategies, utilize critical thinking to make sense of problems and persevere in solving them, and work productively in teams while using cultural global competence to list a few of the NJDOE standards geared to college and career readiness. (NJDOE, 2021)

Subsequently, this study of the 2018 PSAT/NMSQT utilizing the HCRM to evaluate the language of each question complements the research of others in the field of education in which the measurement is not the future long term correlating data of the 2018 PSAT/NMSQT scores, but an accurate, objective, and thorough measurement of how test-takers apply higher-order thinking skills to answer each question. Inherently, the array of characteristics and traits in each of the Revised Bloom's Taxonomy and Webb's Depth of Knowledge (DOK) meticulously organized cells in the subject-specific HCRMs provided the evaluators of the 2018 PSAT/NMSQT with a clearer picture of the higher-order thinking skills required to complete each question successfully. Most importantly, the intentionality in choosing subject-specific HCRMs contributed to the substantive nature of this research through the ability to objectively and precisely score each question's cognitive rigor. In effect, the researchers were able to add to the body of research in the field of education surrounding standardized testing, the assessment of higher-order thinking skills, the preparation for college and career readiness, and what skills were strictly used on each question of the 2018 PSAT/NMSQT by using a valid, reliable, and neutral measurement tool.

Part of the outcome of this study revealed a substantial minority of questions on the 2018 PSAT/NMSQT reaching Level 3 or Level 4 on Webb's Depth of Knowledge axis of the HCRM. Conversely, the assessment and the tabulation of the 2018 PSAT/NMSQT has resulted in 31.6% of the questions scoring in the region of higher-order thinking skills as compared to Hess' Cognitive Rigor Matrix definition of higher-order thinking. The regions of the HCRM that promote Webb's DOK Level 3; Strategic Thinking and Webb's DOK Level 4; Extended Thinking establish the prompt as being a Higher-Order Thinking question. Specifically, in order to assess a student's college and career readiness, according to the subject-specific HCRM, test-takers should be asked to relate mathematical concepts to other content areas, select or devise an approach among many alternatives to solve a problem, explain how concepts or ideas specifically relate to other content domains, illustrate how multiple themes may be interrelated, analyze multiple sources of evidence, or evaluate relevance, accuracy, and completeness of information across multiple sources. Unfortunately, the majority of the questions (68.4%) were scored in the region of Webb's DOK Level 1; Recall or Webb's DOK Level 2; Skills and Concepts, as compared to Hess' Cognitive Rigor Matrix definition of higher-order thinking. Essentially, the majority of the questions prompted students to apply rules of grammar or use resources to edit specific spelling, grammar, punctuation, conventions, or word use, use knowledge to solve routine math problems, specify and explain relationships, or use context to identify the meaning of words or phrases. Notably, the researchers discovered that the majority of the questions on the 2018 PSAT/NMSQT asked students to understand a concept and apply skills and concepts related to that concept. The amount of questions that asked students to connect ideas, make inferences, create, synthesize, or analyze multiple sources of information was in the solid minority.

Significantly, the new knowledge gained as a result of this study can be characterized as supporting and adding to the research of other academics and authors. Specifically, College and Career Readiness Skills detailed by Conley, the transfer skills illustrated by McTighe and Wiggins, and the survival skills highlighted by Wagner are all largely absent on the 2018 PSAT/NMSQT, according to this study's research. A powerful point that the researchers believe is worth emphasizing by Metha and Fine:

> Fundamentally, the study resulted in about four out of five classrooms Mehta and Fine visited featured tasks that were in the bottom half of Bloom's taxonomy, asking students to recall, comprehend, or apply, rather than to analyze, synthesize, or create" (Mehta & Fine, 2011).

One of the quintessential assessments of higher-order thinking skills, the PSAT/NMSQT, fell demonstrably short of expectations and nearly aligns perfectly to the data of Mehta and Fine (2011) in which the majority of the tasks were in the bottom half of Bloom's taxonomy and Webb's Depth of Knowledge.

In comparing this study to other research on the PSAT/NMSQT and education in the United States, the researchers have added to the field of study an objective measure of the language used on the 2018 PSAT/NMSQT and found that the extent to which higher-order thinking were being used on the exam does not support its stated purpose of assessing college readiness. Along these lines, the researchers were able to contextualize and gauge the measurement of higher-order thinking skills on the 2018 PSAT/NMSQT and assess higher-order thinking skills as a whole. By methodically combing through the language of every single question on the 2018 PSAT/NMSQT, the researchers effectively operationalized the HCRM's utility in assessing higher-order thinking skills and cognitive rigor. In essence, a standardized assessment geared toward college and career readiness had yet to be equitably calibrated on such an extensive level. As a result, the independent coders illustrated that higher-order thinking skills were not required to formulate an answer on the 2018 PSAT/NMSQT the solid majority of the time. Accordingly, the long term benefit of the data generated by this study could be intrinsically motivating school leaders, educators, community members, and students to work toward acquiring and practicing higher-order thinking skills, and also developing significantly more effective assessments of higher-order thinking skills by using the HCRM as a guide. Tangentially related to this study is the common practices of educators in the United States school system. The teachers who work with students every single day could read this study and understand that, as a whole, the field of education can increase their proficiency in assessing higher-order thinking skills and cognitive rigor by using the subject-specific Hess' Cognitive Rigor Matrix to create more effective assessments and propel our students to have a stronger grasp of higher-order thinking skills. This study shows that educators in the United States may need to objectively revisit what is asked of students by using the subject-specific HCRMs to help promote higher-order thinking skills for students.

Conclusion

The 2018 PSAT/NMSQT does not contain a majority of questions that compel students to apply higher-order thinking skills as defined by the subject-specific Hess' Cognitive Rigor Matrices. The College Board, the creators of the PSAT/NMSQT and the SAT state that the post-2015, updated versions of the exam "focus on the knowledge, skills, and understandings that research has identified as most important for college and career readiness and success [and have placed] greater emphasis on the meaning of words in extended contexts and on how word choice shapes meaning, tone, and impact" (College Board, 2021). Essentially, the College Board website emphasizes college and career readiness skills being at the forefront of the redesigned PSAT and SAT; one would expect the 2018 PSAT/NMSQT to contain more than 31.6% of the questions that require test-takers to use Webb's DOK Level 3; Strategic Thinking or Webb's DOK Level 4; Extended Thinking. Consequently, from the research performed in this study, the 2018 PSAT/NMSQT has fallen short of an assessment that may measure a test-taker's higher-order thinking skills.

The 2018 PSAT/NMSQT did indeed contain 31.6% of questions that accessed a test-taker's higher-order thinking skills. Compellingly, the Reading Section of the 2018 PSAT/NMSQT was able to break free from the restrictions of a multiple choice style of question, and ask test-takers to use Webb's DOK Level 3, Strategic Thinking. The high concentration (27 out of 47 questions) of Higher-Order Thinking questions, as compared to the other sections, impressively necessitated that students apply word choice, point of view, style to impact readers' or viewers' interpretation of a text, analyze interrelationships among concepts, issues, or problems, use reasoning, planning, and evidence to support information to answer the question successfully, or cite evidence and develop a logical argument for conjectures. In stark contrast, the Math Sections (both calculator and non-calculator sections) asked students to solve routine problems by applying multiple concepts or decision points or retrieve information from a table, graph, or figure and use it to solve a problem requiring multiple steps. Juxtaposing the scarcity of the Level 3 questions in the Math Sections against the abundance of Level 3 questions of the Reading section, one can draw a conclusion that the multiple-choice ecosystem of the 2018 PSAT is capable of asking students to use strategic thinking and reasoning skills, but the breadth of the questions that reach higher-order thinking skill status, as compared to the HCRM, fails to deliver an assessment that measures college and career readiness through the lens of the latest research available. Hence, researchers and authors such as Hess, McTigue, Wiggins, Zhao, Mehta, Fine, Dintersmith, and Wagner, and only highlight the importance of students being able to create, synthesize, design, connect, and critique. In essence, educators have a solemn responsibility to "enable learners to develop significant understanding of core academic content, exhibit critical thinking and problem-solving skills, collaborate, communicate, direct their own learning and possess an academic mindset" (Mehta & Fine, 2019). By the same token, Wagner (2009) defines career and college readiness skills with a list of integral skills necessary for today's students: critical thinking and problem solving, collaboration across networks and leading by influence, agility and adaptability, initiative and entrepreneurialism, effective oral and written communication, accessing and analyzing information, and curiosity and imagination". An even stronger indication of the critical nature higher-order thinking skills plays in the future of students being contributing citizens to the democratic society of the future, they must be equipped with opportunities to sharpen the following attributes according to Conley (2010), an expert on career on college readiness: problem formulation, research, interpretation, communication, precision and accuracy. One can see a pattern emerge that illuminates the shortcomings of the PSAT/NMSQT that mostly asks test-takers to apply recalled knowledge or apply skills on approximately half of the questions (51.8%). Conversely, to echo the point of Wagner and Dintersmith (2015), "we have been conditioned to view some types of credentials as being marks of outsized distinction… Education credentials are our country's caste system." This study has practically adorned and deified the incredibly precious nature of the knowledge and skills that high schools in the United States are responsible for teaching our students to propel them to success in whatever path they choose after graduation.

Substantively, the 2018 PSAT/NMSQT has shown ability, at times, to compel test-takers to access higher-order thinking skills (Webb's DOK; Level 3). However, the 2018 PSAT/NMSQT does not require test-takers to access the highest level (Webb's DOK; Level 4) of higher-order thinking skills a single time. The Reading and the Writing section attempted to push the test-takers to thinking strategically or extend the test-takers' thinking by asking the student to evaluate the validity of evidence or cite evidence or justify or critique conclusions drawn within the confines of a mass-produced, standardized multiple-choice examination. Glaringly, the Math sections fell short in constructing math problems along the lines of relating math concepts to other content areas, asking a student to devise an approach among many alternatives to solve a problem, analyze multiple sources of evidence, gather, analyze, and evaluate information to draw conclusions, or design a mathematical model to inform and solve a practical or abstract solution. Vitally, if the College Board's PSAT/NMSQT aspires to be the premiere measurement tool to assess college and career readiness, 31.6% of the question reached the threshold of higher-order thinking as compared to the subject-specific HCRM is insufficient.

Recommendations of Practice

Distinctly, the body of research within the field of education that addresses standardized testing has not investigated an examination as prominent or widely used as the PSAT through the viewpoint of subject-specific Hess' Cognitive Rigor Matrices. Fundamentally, the researchers in this study of the 2018 PSAT/NMSQT have laboriously established that the majority of the questions do not require the test-taker to employ higher-order thinking skills in order to successfully answer the prompt. Correspondingly, the new evaluative data from this study can have a wide range of implications for a wide array of stakeholders of educational communities. From a satellite view, the distinct impact of the data outcomes of the investigation of the language used in each question PSAT/NMSQT on the culture of public education should be recognized as an affirmation for those who eschew the standardized testing driving the pedagogy that focuses solely on standardized tests to accurately measure growth or development in a student's ability to design, connect, synthesize, or create (Webb's DOK Level 4). Using a far more specific lens of the implications of findings from the research, there are several recommendations for the PSAT/NMSQT to improve upon their frequency of questions that access higher-order thinking skills. The Math Section scored the lowest of the three sections on the 2018 PSAT/NMSQT mostly due to the questions focusing on asking the test-taker to solve routine problems (application of skills and concepts… Coordinate 2, 3). In order for the Math Section to improve upon the frequency in which test-takers access higher-order thinking skills as compared to the language used on the Math CRM, the recommendation is for the PSAT/NMSQT to contain more Math questions that ask students to gather, analyze, and evaluation information, synthesize information across multiple sources or even on esource, analyze multiple sources of evidence, draw conclusions from data while citing evidence, generalize a pattern, interpret data from a complex graph, or verify the reasonableness of results. The Writing Section also contained a minority of prompts that required the test-taker to use higher-order thinking skills to complete the question proficiently. The recommendation for the Writing Section on the PSAT/NMSQT would be to reduce the number of questions that asked the test-taker to apply grammar rules (Coordinate 1,3) and to increase the number of questions that asked students to use strategic reasoning or extended thinking. Sufficiently, 11 out of the 44 questions in the Writing Section of the 2018 PSAT/NMSQT did compel the test-taker to apply strategic thinking (Coordinate 3,3), and the researchers of this study recommend increasing the number of questions that asked test-takers to review a final draft for meaning, progression of ideas, or logic chain or analyze interrelationships among concepts issues, and problems ina test. Furthermore, the Writing Section should also include prompts that require test-takers to use extended thinking skills in which analysis of multiple sources of evidence or numerous works by the same author or across genres is assessed at a higher rate within the Writing Section. Additionally, the researchers would like to highlight the effectiveness of the Reading Section in achieving a scoring rate of higher-order thinking questions the majority of the time within the 2018 PSAT/NMSQT. The multiple choice nature of the PSAT can be challenging to access higher-order thinking skills, but the Reading Section contained questions that reached the highest Coordinate (5,3) out of all the sections. Particularly, the Reading Section asked test-takers to not only understand strategic reasoning at a higher frequency, but the Reading Section also pushed test-takers all the way into the Evaluate Row and Strategic Thinking Column of the Reading CRM asks test-takers to cite evidence and develop a logical argument for conjectures. The constructive feedback for the 2018 PSAT/NMSQT Reading Section would be to have more questions that ask students to synthesize information within one or more texts and to evaluate relevance, accuracy, and completeness of information from multiple sources. Overall, the Reading Section buoyed the results of the 2018 PSAT/NMSQT, and with a refinement of the language on the Math and Writing Section in concordance with the HCRM, the PSAT/NMSQT in the future can be viewed as a more effective tool that measures College and Career Readiness.

The application of the findings from this study can influence students' decisions, the pedagogy of teachers, and the decision-making process of those responsible for local- and state-level curriculum or even the graduation requirements of high school students. Ultimately, high school students and the parents and/or guardians of the teenagers pushing themselves to be evaluated by the PSAT/NMSQT and its far-reaching consequences should be informed that the majority of the questions on the PSAT/NMSQT ask students to simply apply skills. Consequently, stakeholders with an enlightened perspective of the PSAT/NMSQT should also shift their mindset to use the PSAT/NMSQT for what the test really is - a formative assessment that can identify some areas of improvement that may help students be more prepared for the SAT. The claim stating that the PSAT/NMSQT is a measurement for college readiness has fallen flat. Simply put, as a result of this study, the high-stakes of standardized testing needs to be reduced. Consequently, school leaders, administrators, and state-level educational leaders should consider the findings of this study as a cautionary tale, and the emphasis on test scores should be reduced with more focus placed on developing higher-order thinking skills on a daily basis. A student's test score is just one measure, one moment, one scale in which a student's ability to apply concepts is measured. A student's test score, according to the data explored and organized in this study indicating the minority of questions that actually examine higher-order thinking skills, should not be the defining achievement or failure of a student's academic career. Most importantly, asking students to be prepared for college and career readiness should include more opportunities to sharpen their Webb's DOK Level 4 Extended Thinking skills. Local education agencies have the ability to ask more effective questions from and create more effective tasks for their students to practice higher-order thinking skills and ask more of their teachers to help boost learning communities to new heights and higher expectations of growth and development in the regions of strategic thinking and extending thinking. The HCRM is a well-acquitted, thoroughly researched measurement tool that can be applied to local assessments and cultivate a culture of advancement deeper learning by designing assessments, unlike the 2018 PSAT/NMSQT. Thus, a prescriptive, yet differentiated catalog of refined assessments that asks the students to solve real-world, open-ended complex problems that do not have a singular right answer but require a student to research, justify, cite evidence, and design a solution is the most appropriate assessment of high school students' proficiency of higher-order thinking skills. Conclusively, local education agencies have the obligation to design and implement assessments that compel students to illustrate meaningful knowledge acquisition and incorporate authentic opportunities for students to practice higher-order thinking skills as categorized by Hess' Cognitive Rigor Matrices.

Recommendations for Policy

The PSAT/NMSQT should be globally viewed by students, school districts, and state policymakers for what it is: a test that gauges some skills related to math, reading, and writing proficiency. The data from this study indicate the PSAT/NMSQT expectation of a test-taker is to apply skills and concepts or apply recall to successfully complete prompts on the exam. Any other expectation of the test behind application of skills and concepts and a test that prompts students to use strategic or extending thinking (Webb's DOK Level 3 or Level 4) is incorrect. Therefore, reducing the significance of the PSAT/NMSQT as an indicator for career or college readiness, an indicator for a requisite score to unlock courses for a high school student, or even perhaps being considered as a milestone to indicate high school graduation proficiency would be the best course of action for Local Education Agencies and State Policy makers. Unfortunately, the 2018 PSAT/NMSQT did not climb the subject-specific Hess' Cognitive Rigor Matrices scale high enough to earn a distinction of a higher-order thinking skill examination.

State policymakers and Local Education Agencies should be asking themselves if a multiple-choice test that contains questions that compels students to access higher-order thinking skills at a rate of 31.6% worth the time, resources, logistics, disruption to learning, and gravitas that the PSAT/NMSQT currently requires? Since the global pandemic, the test-blind movement in the universe of college admissions has gained momentum.

> A growing number of U.S. colleges and universities are abandoning ACT and SAT scores as part of their admissions process. The list of schools dropping the exams includes Northern Illinois University; Reed College in Oregon; Hampshire College in Massachusetts; Loyola University in New Orleans; the University of New England;

> Washington State University; and some University of California campuses, including Berkeley. Admissions staff at these schools cite two basic reasons for dropping the ACT and SAT, long a rite of passage for high school students planning for college. The first is one of simple scheduling given the pandemic: Canceled testing days in the spring and summer made it impossible for students to take exams in time for fall admissions. The second, however, is more fundamental — some education experts say the ACT and SAT are a poor predictor of whether a student will succeed in college. (Brooks, 2020)

The researchers began this study in early 2018, and the world has rapidly changed over the course of the three years in completing this dissertation. Astoundingly, the data from this study could not be more relevant in relation to the incredible seismic shift in the college admissions landscape, drastically reducing the importance of the SAT or ACT. The data from this study can play a prominent role in determining the future of the college admissions process and how high school students should prepare for post-graduate success. The local education agencies and state policy-makers should be interested in the accuracy, functionality, objectivity, and universal application of the subject-specific HCRM when measuring an assessment that supposedly addresses higher-order thinking skills.

Recommendations for Further Research

Encouragingly, the specific corner of our education field that focuses on effective assessments has shifted to more experiential, practical, and real-world application of knowledge and skills. Additionally, open-ended prompts or assignments that force students to make meaning out of what they learned in the classroom have become paramount. The research conducted within this study found terrain that had yet to be explored. The PSAT/NMSQT had yet to be examined on a question by question basis through the lens of Hess' Cognitive Rigor Matrix in a deep analysis of how the language of each question compelled the student to use or not use higher-order thinking skills. Thus, much like the study of Sydoruk (2018) and Fitzhugh (2019), this research should also be used as a springboard for other explorers of education. The explorers can use the subject-specific Hess' Cognitive Rigor Matrices on a multitude of fascinating assessments that claim to be higher-order thinking purveyors.

The approach to analyzing the PSAT/NMSQT in this study can be used to investigate higher-order thinking skills as defined by the subject-specific Hess' Cognitive Rigor Matrices for a singular assessment or a series of assessments by other researchers. Furthermore, the PSAT/NMSQT has been used for decades to purportedly measure a student's readiness for college. Researchers may want to replicate this study of the 2018 PSAT/NMSQT and conduct a longitudinal study of PSATs across a ten or twenty-year period to reveal the consistency or inconsistency of the PSAT/NSMQT's ability to indicate a student's preparations for college. Also, other researchers may want to look into studying the SAT, the ACT, the GRE, or the LSAT using the HCRM as a scale to identify the frequency or range of questions that assess various higher-order thinking skills. Another interesting angle of further study that would certainly add to the field of educational assessments would be to blindly compare the PSAT/NMSQT's score report that labels each question as easy, medium, or hard to the scoring of each question using the subject-specific Hess' Cognitive Rigor Matrices. Lastly, additional research related to this study that would advance the field of effective educational assessments would be to compare localized assessments from a school district or a collection of school districts to the subject-specific Hess' Cognitive Rigor Matrices in an effort to explore the higher-order thinking skills students are being compelled to use on a yearly basis in the classroom. This particular study could be underpinned by focusing on the higher-order thinking skills assessed across district factor groups, and using socio-economics as a backdrop of the study to reveal how different types of students are given chances to sharpen their higher-order thinking skills.

References

Anderson, L. (2020, October 27). *Taxonomies of educational objectives as bases for curriculum planning.* Oxford Research Encyclopedia of Education. Retrieved 18 Dec. 2020, from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-1038.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Allyn & Bacon. Boston, M.A. (Pearson Education Group).

Anderson, L. W., Sosniak, L. A., Bloom, B. S., & The National Society for the Study of Education. (1994). *Bloom's taxonomy: A forty-year retrospective.* Chicago, I.L.: NSSE.

Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry, 11*(1), 3-42. https://doi.org/10.2307/1179509.

Arneson, J. J., Sackett, P. R., & Beatty, A. S. (2011). Ability-performance relationships in education and employment settings: Critical tests of the more-is-better and the good-enough hypotheses. *Psychological Science, 22*(10), 1336-42. doi: 10.1177/0956797611417004.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* The University of Chicago Press.

Aungst, G. (2014, September 4). *Using Webb's depth of knowledge to increase rigor.* Edutopia. Retrieved from https://www.edutopia.org/blog/webbs-depth-knowledge-increase-rigor-gerald-aungst#:~:text=One%20useful%20tool%2C%20Norman%20Webb%27s,required%20to%20successfully%20complete%20them

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, N.Y.: McKay.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.

Brooks, K. J. (2020, October 7). *ACT and SAT SCORES no longer required for admissions at some colleges.* CBS News. Retrieved from https://www.cbsnews.com/news/act-and-sat-no-longer-required-college-admissions/.

Bowen, G. A. (2009). *Document analysis as a qualitative research method. Qualitative Research Journal, 9*(2), 27-40.

Clark, B. (1983). *Growing up gifted.* Pearson.

College Board. (2018a, March 06). Retrieved from https://collegereadiness.collegeboard.org/sat/inside-the-test/key-features

College Board. (2018b, March 06). Retrieved from https://collegereadiness.collegeboard.org/psat-nmsqt-psat-10/inside-the-test

College Board. (2020). *Overview. About the College Board.* Retrieved from https://about.collegeboard.org/overview

CollegeEd. (n.d.). College Planning Program: CollegeEd is standards based. Retrieved December 31, 2020, from https://bigfuture.collegeboard.org/get-started/educator-resource-center/collegeed-college-planning-program/collegeed-is-standards-based

Conley, D. (2003). *Understanding university success:. A report from Standards for Success: A project of the Association of American Universities and The Pew Charitable Trusts.* Retrieved from https://math.mnrt.net/standards/SandS%20Introduction%20and%20Mathematics.pdf

Conley, D. (2007). *Redefining college readiness.* Educational Policy Improvement Center. Retrieved from https://files.eric.ed.gov/fulltext/ED539251.pdf

Conley, D. (2010). *College and career ready: Helping all students succeed beyond high school.* Jossey-Bass.

Conley, D. (2011). *Four keys to college and career readiness.* Education Policy Task Force Council of State Governments. Retrieved from https://knowledgecenter.csg.org/kc/system/files/conleyPDF.pdf

Corbin, J., & Strauss, A. (2008) *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.

Corwin, Z., & Tierney, W. (2007, January). *Getting there - and beyond: Building a culture of college-going in high schools.* Center for Higher Education Policy Analysis. Retrieved from https://files.eric.ed.gov/fulltext/ED498731.pdf

Edmonds, D. (2015). *Does the PSAT really matter?* Forbes. Retrieved from Rehttps://www.forbes.com/sites/noodleeducation/2015/10/12/does-the-psat-really-matter/?sh=2a39048d5858.

Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice.* Toronto: Collier Macmillan Canada.

Fitzhugh, G. (2019). *A comparison of complex thinking required by the elementary New Jersey student learning standards and past New Jersey curriculum standards* (Doctoral dissertation). Retrieved from Seton Hall University Dissertations and Theses. (2613). https://scholarship.shu.edu/dissertations/2613

Francis, E. M. (2016). *Now that's a good question!: How to promote cognitive rigor through classroom questioning*. Alexandria, V.A.: ASCD.

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

Harlow, D. I. (2019, March 19). *How the ACT and SAT exams are built to fail students.* The Next Web. Retrieved from https://thenextweb.com/podium/2019/03/16/how-the-act-and-sat-exams-are-built-to-fail-students

Henderson, W. D. (2005). The LSAT, law school exams and meritocracy: The surprising and undertheorized role of test-taking speed. *Texas Law Review, 82*(2004), IU Law-Bloomington Research Paper No. 20. Retrieved from Henderson https://ssrn.com/abstract=465381

Hersh, R. H., & Keeling, R. P. (2011). *We're Losing Our Minds: Rethinking American Higher Education*. Palgrave Macmillan.

Hess, K. (2009, updated 2013). *Linking research with practice: A local assessment toolkit to guide school leaders.* Retrieved from https://www.karin-hess.com/free-resources

Hess, K. (2014). *Where did the Hess CRM come from? CRM History.* Retrieved from https://www.karin-hess.com/crm-history

Hess, K. (2018). *A local assessment toolkit to promote deeper learning: Transforming research into practice*. Thousand Oaks, CA: Corwin

Hess, K., Colby, R., & Joseph, D. (2020). *Deeper competency-based learning: Making equitable, student-centered, sustainable shifts.* Corwin.

Hess, K., Gong, B., & Steinitz, R. (2014, March). *Ready for college and career? Achieving the common core standards and beyond through deeper, student centered learning.* Nellie Mae Education Foundation. Retrieved from https://www.nmefoundation.org/wp-content/uploads/2020/05/Ready-for-College-and-Career.pdf

Hess, K., Jones, B., Carlock, D., & Walkup, J. (2009). *Cognitive rigor: Blending the strengths of Bloom's taxonomy and Webb's depth of knowledge to enhance classroom-level processes.* Online submission. Retrieved from https://files.eric.ed.gov/fulltext/ED517804.pdf

Howard, C. (2014, March 07). *Right answer for SAT? College Board revamps test, partners with Khan Academy.* Forbes. Retrieved from https://www.forbes.com/sites/carolinehoward/2014/03/05/right-answer-for-sat-college-board-revamps-test-adds-partnership-with-khan-academy/?sh=2b7aba1e5d5f

Inflexion. (2020). *Who we are.* Retrieved from https://www.inflexion.org/who-we-are/history/

Iowa State University. (2020, November 04). *Home.* Retrieved from https://www.celt.iastate.edu/

Iowa State University Center for Excellence in Learning and Teaching. (2020). *Revised Bloom's Taxonomy.* Retrieved from https://www.celt.iastate.edu/teaching/effective-teaching-practices/revised-blooms-taxonomy/

Kastberg, S. (2003). Using Bloom's Taxonomy as a framework for classroom assessment. *Mathematics Teacher. 96*(6), 402-405.

Kelly, P., & Ewell, P. (2009). *Colleges and universities and their stewardship of place: A guide for developing performance measures for the equity of access and student success.* National Center for Higher Education Management Systems. Retrieved from https://files.eric.ed.gov/fulltext/ED512583.pdf

Kobrin, J., Patterson, B., Shaw, E., Mattern, K., & Barbuti, S. (2008). Validity of the SAT® for predicting first-year college grade point average. *College Board Research Report, 2008*(5). Retrieved from https://files.eric.ed.gov/fulltext/ED563202.pdf

Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. *Teachers College Record, 95*(6), 64-81.

Law School Admission Council. (2020). *The LSAT.* Retrieved from https://www.lsac.org/lsat

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice, 32*(3), 131-137. Retrieved November 10, 2020, from http://www.jstor.org/stable/1476693

Linn, R., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*(1), 29-36. https://doi.org/10.3102%2F01623737024001029

Lombardi, A. R., Kowitt, J. S., & Staples, F. E. (2015). Correlates of critical thinking and college and career readiness for students with and without disabilities. *Career Development and Transition for Exceptional Individuals, 38*(3), 142–151. https://doi.org/10.1177/2165143414534888

Marini, J., Mattern, K., & Shaw, E. (2011). Examining the linearity of the PSAT/NMSQT® –FYGPA relationship. C*ollege Board Research Report, 2011*(7). Retrieved from https://files.eric.ed.gov/fulltext/ED562677.pdf

Marsh, H., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs? *The Journal of Higher Education, 73*(5), 603-641. Retrieved December 18, 2020, from http://www.jstor.org/stable/1558435

Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research, 1*(2). Retrieved March 10, 2005 from http://www.qualitative-research.net/fqs-texte/2-00/02-00mayring-e.htm

Massachusetts Institute of Technology. (n.d.). *Admissions Statistics.* MIT Admissions. Retrieved November 12, 2020, from https://mitadmissions.org/apply/process/stats/

McTighe, J., & Silver, H. F. (2020). *Teaching for deeper learning: Tools to engage students in meaning making*. Alexandria, V.A.: ASCD.

Mehta, J., & Fine, S. (2019). *In search of deeper learning: The quest to remake the American high school*. Cambridge, M.A.: Harvard University Press.

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

New Jersey Department of Education. (2016). *N.J.A.C. 6A:8, STANDARDS AND ASSESSMENT*. https://www.state.nj.us/education/code/current/title6a/chap8.pdf

New Jersey Department of Education. (2012, April 30). *Final report of the New Jersey Department of Education College and Career Readiness Task Force.* https://www.state.nj.us/education/genfo/readiness.pdf

New Jersey Department of Education. (2019). Retrieved from https://rc.doe.state.nj.us/SearchForSchool.aspx

Niebling, B. C. (2012). *Determining the cognitive complexity of the Iowa Core in Literacy and Mathematics: Implications and applications for curriculum alignment. Iowa Department of Education.* Des Moines, IA: Retrieved from http://www.educateiowa.gov/index.php?option=com_docman&task=doc_download&gid=14524&Itemid=5111

Pellegrino, J. W. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Washington, D.C.: National Academies Press.

Peterson, J. (2017). *The surprising power of the PSAT.* HuffPost. Retrieved from https://www.huffpost.com/entry/the-surprising-power-of-t_b_8113822.

Pink, D. (2005). *A whole new mind: Why right-brainers will rule the future*. The Berkley Publishing Group.

Portrait of a Graduate by Battelle for Kids (2019) *Do you have a Portrait of a Graduate?* Retrieved from https://portraitofagraduate.org/

Proctor, T., Wyatt, J., & Wiley, A. (2010). *The development of early indicators of college readiness using the PSAT/ NMSQT and ReadiStep examination programs* (College Board Research Report). New York, N.Y.: The College Board.

Roediger, H., & Marsh, E. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5). Retrieved from http://psychnet.wustl.edu/memory/wp-content/uploads/2018/04/Roediger-Marsh-2005_JEPLMC.pdf

Rodriguez, P. (2020). *The SAT is helping minorities succeed.* Capitol Weekly. Retrieved from https://capitolweekly.net/the-sat-is-helping-minorities-succeed/

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist, 59*(1), 7–13. https://doi.org/10.1037/0003-066X.59.1.7

Schlager, K. (2018, October 9). *The top New Jersey public high schools 2018*. New Jersey Monthly. Retrieved from https://njmonthly.com/articles/towns-schools/the-top-new-jersey-public-high-schools-2018/

Schulz, H. W., & Fitzpatrick, B. (2015, November 30). *Teachers' understandings of critical and higher order thinking and what this means for their teaching and assessments.* Retrieved from https://eric.ed.gov/?id=EJ1113229

Sforza, D. (2014). The inclusion of cognitive complexity: A content analysis of New Jersey's current and past intended curriculum (Doctoral dissertation). Retrieved from Seton Hall eRepository http://scholarship.shu.edu/dissertations/2020

Simpson, A., Mokalled, S., Ellenberg, L. A., & Chu, S. M. (2014). A Tool for Rethinking Teacher's Questioning. *Mathematics Teaching in the Middle School, 20*(5), 294-302. https://www.jstor.org/stable/10.5951/mathteacmiddscho.20.5.0294

Sizer, T. R. (1984). *Horace's compromise: The dilemma of American high schools.* Houghton Mifflin.

Soares, L. (2009). *Working learners educating our entire workforce for success in the 21st century.* Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/issues/2009/06/pdf/working_learners.pdf?_ga=2.37079264.1700875039.1605320430-2072692498.1605320430

Strauss, V. (2019, April 18). *Can coaching truly boost SAT scores? For years, the College Board said no. Now it says yes.* Washington Post. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2017/05/09/can-coaching-truly-boost-sat-scores-for-years-the-college-board-said-no-now-it-says-yes/

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

The Glossary of Education Reform. (2014, March 5). *Bloom's Taxonomy.* Retrieved from https://www.edglossary.org/blooms-taxonomy/

The Glossary of Education Reform. (2015, November 12). *Standardized Test.* Retrieved from https://www.edglossary.org/standardized-test/

The Standards Company LLC. (2008a). “Study of the alignment of student assignments to the academic standards in the state of Nevada pursuant to Senate Bill 184, Chap. 420, Statutes of Nevada 2007." Retrieved September 28, 2021, from Legislative Counsel Bureau, Nevada State Legislature, technical report, http://www.leg.state.nv.us/lcb/fiscal/Final_Report-Curriculum_Study.pdf.

Tienken, C. (2008). Rankings of International Achievement Test performance and economic strength: Correlation or conjecture?. *International Journal of Education Policy and Leadership, 3.*

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

Tigas, M. (2013, May 09). *College Board - Nonprofit Explorer.* Retrieved from https://projects.propublica.org/nonprofits/organizations/131623965

University of California Academic Senate. (2020, February 2). *Report of the UC Academic Council Standardized Testing Task Force.* Retrieved April 1, 2020, from https://senate.universityofcalifornia.edu/_files/underreview/sttf-report.pdf

View, J. L. (2012). Trapped by bubble sheets: Teachers struggling to deepen their historical knowledge. *Transformations, 23*(1), 58-75. Retrieved November 14, 2020, from http://www.jstor.org/stable/10.5325/trajincschped.23.1.0058

Vincent, W. (2018, January 25). *The evolving online test prep market.* Retrieved from https://wsvincent.com/the-evolving-online-test-prep-market/

Wagner, T. (2009). *Global achievement gap*. New York, NY: Basic Books.

Wagner, T., & Dintersmith, T. (2015). *Most likely to succeed: Preparing our kids for the innovation era*. New York, NY: Scribner.

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

Wolfe, K. E. (2004). *Understanding the careers of the alumni of the MIT Mechanical Engineering Department.* Retrieved from https://dspace.mit.edu/handle/1721.1/32796

Wyatt, J. N., Kobrin, J. L., Wiley, A., Camara, W., & Proestler, N. (2011). *SAT benchmarks: development of a college readiness benchmark and its relationship to secondary and postsecondary school performance. Research Report 2011-5.* College Board. Retrieved from https://eric.ed.gov/?id=ED521173

Wiley, A., Wyatt, J., & Camara, W. (2010). *The development of a multidimensional college readiness index*. College Board. Retrived from https://www.newenglandssc.org/wp-content/uploads/2015/11/collegeboard_sat_collegereadinessindex.pdf

Yuan, K., & Le, V.-N. (2012). *Measuring deeper learning through cognitively demanding test items.* Rand Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR483.html

Zenko, J. (accessed 2021). *When (and why) to use heat maps.* Enterprise BI, Reporting & Analytics Software. Retrieved from https://www.dundas.com/resources/blogs/best-practices/when-and-why-to-use-heat-maps#:~:text=By%20definition%2C%20Heat%20Maps%20are,re%20much%20more%20than%20that.

Zhao, Y. (2012). *World class learners: Educating creative and entrepreneurial students* (1st ed.). Corwin.

Appendix

Appendix A: Hess Cognitive Rigor Matrix: Close Reading and Listening, (Hess, K., 2009, updated 2013)

TOOL 1

HESS COGNITIVE RIGOR MATRIX | READING-LISTENING CRM

Integrating Depth-of-Knowledge Levels with Bloom's Cognitive Process Dimensions

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

Appendix B: Hess Cognitive Rigor Matrix: Math and Science, (Hess, K., 2009, updated 2013)

TOOL 2

HESS COGNITIVE RIGOR MATRIX | MATH-SCIENCE CRM

Integrating Depth-of-Knowledge Levels with Bloom's Cognitive Process Dimensions

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

Appendix C: Hess Cognitive Rigor Matrix: Written and Oral Communication, (Hess, K., 2009, updated 2013)

TOOL 3

HESS COGNITIVE RIGOR MATRIX | WRITING-SPEAKING CRM

Integrating Depth-of-Knowledge Levels with Bloom's Cognitive Process Dimensions

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

Appendix D: Coding Table

Section	Question #	Type of Question	Webb Axis Score	Bloom Axis Score	Overall score
Reading	1	Multiple Choice	2	2	2, 2
Reading	2	Multiple Choice	3	2	3,2
Reading	3	Multiple Choice	2	2	2, 2
Reading	4	Multiple Choice	3	2	3,2
Reading	5	Multiple Choice	2	3	2,3
Reading	6	Multiple Choice	1	2	1,2
Reading	7	Multiple Choice	3	2	3,2
Reading	8	Multiple Choice	3	2	3,2
Reading	9	Multiple Choice	3	3	3,3
Reading	10	Multiple Choice	3	4	3,4
Reading	11	Multiple Choice	3	2	3,2
Reading	12	Multiple Choice	2	3	2,3
Reading	13	Multiple Choice	3	4	3,4
Reading	14	Multiple Choice	3	2	3,2
Reading	15	Multiple Choice	3	3	3,3
Reading	16	Multiple Choice	2	3	3,2

Reading	17	Multiple Choice	2	2	2,2
Reading	18	Multiple Choice	3	4	3,4
Reading	19	Multiple Choice	1	4	1,4
Reading	20	Multiple Choice	3	3	3,3
Reading	21	Multiple Choice	3	2	3,2
Reading	22	Multiple Choice	3	2	3,2
Reading	23	Multiple Choice	2	3	2,3
Reading	24	Multiple Choice	2	3	2,3
Reading	25	Multiple Choice	2	2	2, 2
Reading	26	Multiple Choice	3	2	3,2
Reading	27	Multiple Choice	1	4	1,4
Reading	28	Multiple Choice	3	4	3,4
Reading	29	Multiple Choice	3	3	3,3
Reading	30	Multiple Choice	3	2	3,2
Reading	31	Multiple Choice	2	3	2,3
Reading	32	Multiple Choice	3	3	3,3
Reading	33	Multiple Choice	3	2	3,2
Reading	34	Multiple Choice	2	3	2,3

Reading	35	Multiple Choice	3	3	3,3
Reading	36	Multiple Choice	1	1	1,1
Reading	37	Multiple Choice	1	1	1,1
Reading	38	Multiple Choice	3	3	3,3
Reading	39	Multiple Choice	2	3	2,3
Reading	40	Multiple Choice	2	3	2,3
Reading	41	Multiple Choice	3	2	3,2
Reading	42	Multiple Choice	3	2	3,2
Reading	43	Multiple Choice	1	4	1,4
Reading	44	Multiple Choice	3	4	3,4
Reading	45	Multiple Choice	2	4	2,4
Reading	46	Multiple Choice	3	5	3,5
Reading	47	Multiple Choice	3	2	3,2
Writing	1	Multiple Choice	1	2	1,2
Writing	2	Multiple Choice	1	2	1,2
Writing	3	Multiple Choice	2	4	2,3
Writing	4	Multiple Choice	1	3	3
Writing	5	Multiple Choice	3	3	3,3

Writing	6	Multiple Choice	1	3	1,3
Writing	7	Multiple Choice	1	3	1,3
Writing	8	Multiple Choice	3	3	3,3
Writing	9	Multiple Choice	1	3	1,3
Writing	10	Multiple Choice	1	3	1,3
Writing	11	Multiple Choice	1	3	1,3
Writing	12	Multiple Choice	3	3	3,3
Writing	13	Multiple Choice	3	3	3,3
Writing	14	Multiple Choice	1	3	1,3
Writing	15	Multiple Choice	1	3	1,3
Writing	16	Multiple Choice	1	3	1,3
Writing	17	Multiple Choice	1	3	1,3
Writing	18	Multiple Choice	1	3	1,3
Writing	19	Multiple Choice	1	3	1,3
Writing	20	Multiple Choice	3	3	3,3
Writing	21	Multiple Choice	3	3	3,3
Writing	22	Multiple Choice	1	3	1,3
Writing	23	Multiple Choice	1	3	1,3

Writing	24	Multiple Choice	1	3	1,3
Writing	25	Multiple Choice	1	3	1,3
Writing	26	Multiple Choice	1	3	1,3
Writing	27	Multiple Choice	3	3	3,3
Writing	28	Multiple Choice	1	3	1,3
Writing	29	Multiple Choice	3	3	3,3
Writing	30	Multiple Choice	1	3	1,3
Writing	31	Multiple Choice	1	3	1,3
Writing	32	Multiple Choice	1	3	1,3
Writing	33	Multiple Choice	3	3	3,3
Writing	34	Multiple Choice	1	3	1,3
Writing	35	Multiple Choice	1	3	1,3
Writing	36	Multiple Choice	1	3	1,3
Writing	37	Multiple Choice	3	3	3,3
Writing	38	Multiple Choice	1	3	1,3
Writing	39	Multiple Choice	1	3	1,3
Writing	40	Multiple Choice	1	3	1,3
Writing	41	Multiple Choice	3	3	3,3

Writing	42	Multiple Choice	3	4	3,4
Writing	43	Multiple Choice	3	5	3,5
Writing	44	Multiple Choice	1	3	1,3
MathNoCalc	1	Multiple Choice	1	3	1,3
MathNoCalc	2	Multiple Choice	1	3	1,3
MathNoCalc	3	Multiple Choice	2	3	2,3
MathNoCalc	4	Multiple Choice	2	2	2,2
MathNoCalc	5	Multiple Choice	1	4	1,4
MathNoCalc	6	Multiple Choice	2	3	2,3
MathNoCalc	7	Multiple Choice	2	3	2,3
MathNoCalc	8	Multiple Choice	2	3	2,3
MathNoCalc	9	Multiple Choice	2	3	2,3
MathNoCalc	10	Multiple Choice	2	3	2,3
MathNoCalc	11	Multiple Choice	3	3	3,3
MathNoCalc	12	Multiple Choice	2	3	2,3
MathNoCalc	13	Multiple Choice	2	3	2,3
MathNoCalc	14	Free-Response	2	3	2,3
MathNoCalc	15	Free-Response	2	3	2,3

MathNoCalc	16	Free-Response	2	3	2,3
MathNoCalc	17	Free-Response	3	3	3,3
MathCalc	1	Multiple Choice	1	2	1,2
MathCalc	2	Multiple Choice	1	3	1,3
MathCalc	3	Multiple Choice	1	3	1,3
MathCalc	4	Multiple Choice	2	3	2,3
MathCalc	5	Multiple Choice	2	3	2,3
MathCalc	6	Multiple Choice	2	3	2,3
MathCalc	7	Multiple Choice	2	3	2,3
MathCalc	8	Multiple Choice	2	3	2,3
MathCalc	9	Multiple Choice	2	3	2,3
MathCalc	10	Multiple Choice	2	3	2,3
MathCalc	11	Multiple Choice	2	3	2,3
MathCalc	12	Multiple Choice	2	3	2,3
MathCalc	13	Multiple Choice	2	2	2,2
MathCalc	14	Multiple Choice	2	3	2,3
MathCalc	15	Multiple Choice	2	3	2,3
MathCalc	16	Multiple Choice	1	3	1,3

MathCalc	17	Multiple Choice	2	3	2,3
MathCalc	18	Multiple Choice	2	4	2,4
MathCalc	19	Multiple Choice	2	4	2,4
MathCalc	20	Multiple Choice	2	3	2,3
MathCalc	21	Multiple Choice	2	2	2,2
MathCalc	22	Multiple Choice	2	3	2,3
MathCalc	23	Multiple Choice	3	3	3,3
MathCalc	24	Multiple Choice	1	3	1,3
MathCalc	25	Multiple Choice	2	3	2,3
MathCalc	26	Multiple Choice	2	3	2,3
MathCalc	27	Multiple Choice	3	2	3,2
MathCalc	28	Open-Ended	2	3	2,3
MathCalc	29	Open-Ended	2	3	2,3
MathCalc	30	Open-Ended	1	4	1,4
MathCalc	31	Open-Ended	1	2	1,2