

2017

A comparison of self-perceived test-taking motivation on large scale high- and low-stakes tests

Randall Joseph King
king226@live.marshall.edu

Follow this and additional works at: <https://mds.marshall.edu/etd>

 Part of the [Curriculum and Instruction Commons](#), and the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

King, Randall Joseph, "A comparison of self-perceived test-taking motivation on large scale high- and low-stakes tests" (2017). *Theses, Dissertations and Capstones*. 1191.
<https://mds.marshall.edu/etd/1191>

This Dissertation is brought to you for free and open access by Marshall Digital Scholar. It has been accepted for inclusion in Theses, Dissertations and Capstones by an authorized administrator of Marshall Digital Scholar. For more information, please contact zhangj@marshall.edu, beachgr@marshall.edu.

**A COMPARISON OF SELF-PERCEIVED TEST-TAKING MOTIVATION ON LARGE
SCALE HIGH- AND LOW-STAKES TESTS**

Randall Joseph King
Marshall University
College of Education
and Professional Development

Dissertation submitted to the
Graduate College of Marshall University
in partial fulfillment of the
requirements for the degree of

Doctor of Education
in
Curriculum and Instruction

Committee Chair, Dr. Samuel Securro, Jr.
Dr. Lisa A. Heaton
Dr. Ronald Childress
Dr. Paula Nelson

Marshall University
December 2017

ACKNOWLEDGMENTS

I would like to acknowledge my wife, Bianca and my daughters, Samara and Giana for their support during the last decade as I plodded through the doc program at MUGC. Their patience over the years and understanding as I missed out on various activities and milestones has gone above and beyond what anyone has the right to expect. During certain times when my drive was lacking, Bianca was there to push me forward, and I appreciate that we are still married by the end of this.

In addition, I would like to acknowledge my “in-laws,” Sam and Carol Moscar for their continued support and stimulating discussions about my project year after year. Support and accolades from other family members have been much appreciated as well. A very specific “thank you” goes to Mark McMillion for monitoring my progress during a certain time when action was needed.

And many thanks go to my Chair, Dr. Samuel Securo, for his patience, guidance, encouragement, and even his “tough love” when it was necessary.

I also have much appreciation for the members of the 2007 cohort – our time together as we started this journey was definitely memorable. And the food! Within that cohort our little group, Jenny, Hannah, and Kristy, thank you all for the academic challenge and comradery. And specifically to Jenny, you really get to know someone when you carpool every month for four-and-a-half years. I know what you stand for, and I admire you for all that you have accomplished so far. And thank you again for attending my defense.

I also want to express my appreciation to Mike for nearly a lifetime of friendship and for patiently awaiting chess moves while I worked on my paper, to The Huge One for motivational guidance and life coaching, and to Davis Power & the gang for understanding as the fourth horseman faded quietly from the Iron Game.

I especially need to acknowledge Kathy Marino, my math education mentor, who has helped guide me from my high school years through my professional years. I commemorate the day in 1990, when I witnessed my teacher derive the quadratic formula, as the day that sealed my fate as a math educator and amateur mathematician. Her support during my doctoral journey has been invaluable. Many happy Pi Days to you, Mrs. Marino!

And finally, much gratitude to the rest of my Liberty High friends and colleagues – I have appreciated each letter of recommendation, every checkup on my progress, all the Big Franks Fridays, every class covered while I administered my survey, and every pat on the back along the way.

Now to pay it forward

TABLE OF CONTENTS

Acknowledgments.....	ii
Table of contents.....	iii
List of tables.....	vi
Abstract.....	vii
Chapter One: Introduction	1
Background.....	1
Statement of the Problem.....	4
Significance.....	5
Research Questions.....	6
Operational definitions.....	7
Methods.....	8
Delimitations.....	10
Summary.....	10
Chapter Two: Review of Literature	12
Historical Context of K-12 School Testing	12
Increasing Year-End Test Score Performance.....	25
Student Motivation for and Interest in Test Achievement.....	27
Measuring Student Motivation for Test Taking.....	44
Summary.....	51

Chapter Three: Methods	53
Purpose.....	53
Study Design.....	53
Population and Sampling	54
Procedures.....	56
Instrument and Data Collection	58
Survey Instrument.....	60
Data Analysis	61
Summary	63
Chapter Four: Data Analysis and Results	65
Introduction.....	65
Purpose.....	65
Population and Sample	65
Data Analysis and Results	66
Summary.....	75
Chapter Five: Findings, Conclusions and Discussion.....	78
Statement of the Problem.....	78
Research Questions.....	78
Sample and Data Collection.....	79
Conclusions.....	80

Discussion and Implications	85
Recommendations for Further Research.....	90
References.....	93
Appendices.....	107
Appendix A: IRB Approval Letter	107
Appendix B: District-Level Permission.....	108
Appendix C: School Level Permissions	109
Appendix D: Initial Teacher Invitation.....	110
Appendix E: Teacher Instructions	112
Appendix F: Parental Consent Letter.....	113
Appendix G: Student Assent Form.....	115
Appendix H: Student Opinion Survey.....	116
Appendix I: Curriculum Vita.....	119

LIST OF TABLES

Table 1 Harrison County high school enrollments by grade level (2016-2017)..... 55

Table 2 Descriptives for Level of Importance and Effort on State and Admission tests.. 66

Table 3 SOS Item Statistics for State Tests 67

Table 4 SOS Item Statistics for College Tests..... 68

Table 5 Descriptives for future plan, Effort, and Importance for State and College tests 70

Table 6 Descriptive Data for Grade Levels, Type of Tests, Effort, and Importance..... 73

Table 7 Descriptive Data for Gender by Effort and Importance on State Test..... 74

Table 8 Descriptive data for Gender by Effort and Importance on College Tests. 75

ABSTRACT

The central purpose of this study was to investigate the differences in test-taking motivation between large-scale high-stakes and low-stakes tests. The high-stakes test in this study was any college admission test, and the low-stakes test was the mandatory state assessment in West Virginia. Comparisons deduced the influence of the following independent variables on self-reported test-taking motivation: test stakes (high or low), future plan (college-bound or non-college bound), grade level (9 – 12), and sex (female or male).

The sample in this investigation (n=161) was taken from a West Virginia school district with a high school population of about 3,000 students. Participation resulted from random sampling of homeroom teachers and the return of affirmative parental consent forms. Although all subgroups were not represented proportionally, the data were aggregated for analysis.

The dataset for the analysis was the result of voluntary participation in a paper-and-pencil version of the Student Opinion Scale. On this survey, students self-rated their perception of test effort and test importance regarding the high- and low-stakes tests.

The analysis revealed drastically higher motivation from students taking the high-stakes tests compared to students taking the low-stakes test. A similar dramatic difference resulted from comparing the motivation of the college-bound subgroup on the high- and low-stakes tests. Further analyses yielded higher motivation from females and underclassmen when taking the low-stakes state exam.

The central conclusion is that the sample of high school participants in this study indicated higher motivation when taking high-stakes tests compared to low-stakes tests. An important implication is that confidence in state test score validity may be questionable because test scores may not be optimal. Further study is necessary to investigate the scale of this effect.

CHAPTER ONE: INTRODUCTION

Background

Since the inception of No Child Left Behind (NCLB), school systems have felt the increased weight of high-stakes tests. NCLB provides increasingly severe sanctions for consistent low performance and little reward for high performance (U.S. Department of Education, 2010). No education leaders wish for their district or state to be viewed negatively due to poor achievement test scores or low rankings, and consequently, there is great pressure exerted downward through the “educational hierarchy” (Crocco & Costigan, 2007, p 525). As a further result, school districts nationwide attempt many types of fixes to boost test scores, ranging from short-term incentives and cheerleading to long-term adjustment of curriculum and classroom assessment (McColskey & McNunn, 2000).

But on test day, it is neither the state nor the district that will perform well or poorly. Nor is it the school administrator or teacher. The ultimate responsibility lies in the hands of the students. The answer choices that students make on tests determine repercussions or recognition for their schools and districts. Therefore, it seems quite plausible that in addition to routine test preparations, a student’s mindset on test day could have a significant effect on test performance. For example, does the student approach the test seriously, expecting to succeed, or nonchalantly? And to the student, how valuable are the test results?

For optimal test performance and accurate, reliable interpretation of test scores, it seems reasonable that the students must “buy into” the test -- there must be some reward worth the effort.

Mandatory state achievement test

Because the purpose of the state test is to measure student achievement for the benefit of all stakeholders, it can be asked if all the stakeholders even care to know about student achievement. In particular, do all students value this measurement of their own academic achievement? If there were some benefit tied to its test achievement scores, then there may be a value added to the measurement and a resulting increase in effort spent on the test (O’Neil, Sugrue, & Baker, 1995/1996; Sundre & Kitsantas, 2004; Wolf & Smith, 1995). But aside from any student rewards that may exist in individual schools, there is no statewide reward for this achievement score, nor is there any consequence beyond remedial course taking for low test performance. From this viewpoint, the state test has little stakes for students -- the associated importance of “high stakes” is reserved for the educators, school systems, and politicians, including state and national legislators (McGuinness, 2010).

Beyond a potential lack of positive attitudes, certain cultural concerns may actually contribute to negative attitudes toward testing held by parents which are then easily transferred to children. For example, Woodrum’s (2004) interviews of Appalachian parents revealed evidence that cultural values may resist mandatory school testing. The majority of poor Appalachian parents interviewed “found little or no value in state-mandated testing of their children” (p. 5), and many participants responded that they had told their children not to worry about the test or even had kept a child home on test day. In the current study, this cultural concern is quite valid because the sample comes from a West Virginia county school district with about 51% low income students (West Virginia Department of Education, 2012/2013). According to the Appalachian Regional Commission (2015), the economic conditions in Harrison County place it in the middle 50% of U.S. county economies.

College admission tests

Contrast these viewpoints with the analogous view of the ACT or SAT. These tests measure achievement, but not for the benefit of stakeholders; their purpose is to estimate academic readiness for college. The contrast continues as these college admission tests are not free to students or parents; there are monetary costs related to the registration process, and there are limited test administrations. Students do not simply show up to school on a school day and take these tests. They must carry out the process that allows them the opportunity to take them. Even further, there are acceptable levels of ACT and SAT scores for admission into different types of colleges. If students score well enough, they may receive scholarship money to help pay for college. And if they score highly enough, they may even apply to more selective colleges. In contrast, if students score too lowly, then they will have to re-register and pay again to retake the test in order to apply for college. The college admission tests truly have high stakes and direct benefits and consequences for the test-taker.

High-stakes?

For these examples of such high-stakes tests, we must consider upon which stakeholders lay the “high” stakes. In this discussion, it is clear that the state-mandated achievement test has low stakes for students, and the college admission tests have high stakes for students who take those tests. Furthermore, it is clear that the college system presents an opportunity for students to value the outcome of the admission tests whereas the West Virginia Department of Education does not. Again, does the test-taker approach the test seriously, or casually, based on the value of the outcome? Specifically, do state test takers approach the test as seriously as college admission test takers? It may be debated whether students will put forth their best effort on either test, on average, but regardless of the difficulty of such debate, the question of testing

motivation should not remain open-ended; there must be a conclusion. To aggregate achievement test data and base high-stakes decisions on results, the decision makers should take the steps to induce optimal test performance so they can be confident that the data actually do reflect this type of performance. Therefore, the general problem addressed in the current study is that of confidently interpreting test scores on behalf of both the public and the policy makers. Finally, although score interpretation encompasses many large areas such as curriculum, school and district administration, and technical aspects of the tests, these areas will not be approached in this study.

Theoretical framework

The theoretical framework which undergirds this study is the Expectancy-Value model of achievement motivation as developed by Eccles et al. (1983). The theory, in essence, suggests that motivation to achieve a given goal is influenced by the level at which a person *expects* to achieve and by the *value* that is attached to achieving that goal. The combination of these factors results in the effort and attention given to the task. Test-taking motivation is based in the value portion of expectancy-value theory (Sundre, 2007). In particular, testing effort and importance of the test may be considered identical to the cost and attainment value aspects of subjective task value, according to the Eccles et al. (1983) model of expectancy-value theory.

Statement of the Problem

The purpose of this study is to determine what differences exist in test-taking motivation among high school students for completing a college admission test and the annual WV state-mandated achievement test. In the case of the admission test, the exam is not required but has future college going implications for students, including academic scholarships. The opposite is true for the state test, which is required by WV State Policy 2340 (§126-14-4.1), and searches of

best practices have not produced any evidence of personal value for students beyond intrinsic value. However, state testing does have serious consequences for schools and school personnel in the No Child Left Behind compliance malaise. At issue here is whether test takers are giving the same serious and genuine attention to these two different types of academic measure.

Therefore, significant questions are raised: “Are high school students differing in the levels of motivation and energy when taking these exams”?, “ Are these students attaining an optimal score in each case and obtaining results that reflect a true level of school achievement, particularly in the case of the state test? Further, “What differences might exist in testing motivation between college bound and non-college bound students, the sex of the students, and their grade levels?”.

Significance

The level of test-taking motivation on large-scale tests would be valuable information for schools and school systems for both test preparation and interpretation of test results. Decision makers at each level of school administration need to have high levels of confidence in the results of state-mandated tests. Do these scores represent, on average, a valid assessment of student learning and growth, which ultimately can impact curriculum change, teacher evaluation, and legislative actions? At the school level, teachers and administrators are also concerned that the results are valid when being applied to school compliance policy and related probationary rules. As for the college admission tests, colleges and postsecondary programs need to know that the scores represent a reliable level of academic readiness to achieve success at their respective institutions. Also, parents, who are likely paying the costs, need to know that their children are seriously motivated when taking these tests, and if not, they need to know what might be done to increase their engagement.

The concept of a high- and a low-stakes test is chosen for the current study partly because the researcher has several years of test preparation experience in the high school classroom, and partly because such a comparison is both simple and significant. The result of the overall comparison has only two possible outcomes, either of which would be valuable to the pillar of education in West Virginia. One possibility is that there is no statistical difference in the levels of effort students expend on high- and low-stakes tests. Such a result could reinforce confidence in the testing practices in the state if levels are satisfactory or it could encourage change if levels are unacceptable. The other major possibility is a real difference in effort, which could reinforce confidence in current practices for the test. In all, these results can contribute to a growing knowledge base on school standardized testing but specifically address a gap in the knowledge base about the seriousness of student engagement and optimal effort given on mandatory tests across several grade levels.

Statement of Purpose

Therefore, the central purpose of the investigation was to determine the test-taking motivation among high school students in regard to the levels of effort given and importance attached to “low and high stakes” testing. An additional purpose was to determine to what extent test-taking motivation was distinguished by students’ future college plans, grade levels, and sex.

Research Questions

1. What are the levels of test-taking motivation (TTM) reported by high school students who complete the low stakes WV mandatory assessment and the high stakes ACT or SAT college admission tests?
2. What differences exist in TTM reported by high school students who complete the WV mandatory assessment and the ACT or SAT college admission tests?

3. What differences exist in TTM between college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?
4. What are the differences in TTM reported by college-bound high school students who complete the low-stakes WV mandatory assessment and by the same students who complete the high-stakes ACT or SAT college admission tests?
5. What grade level differences exist in TTM reported by students who complete the low-stakes WV mandatory assessment?
6. What grade level differences exist in TTM reported by college-bound high school students who take the high-stakes ACT or SAT college admission tests?
7. What differences exist in TTM reported by girls and boys who complete the low-stakes WV mandatory assessment?
8. What differences exist in TTM between those reported by girls and boys who take the high-stakes ACT or SAT college admission tests?

Operational definitions

1. Test-taking motivation (TTM): in this study, “TTM” referred to the outcomes of the Student Opinion Scale (Sundre & Moore, 2002), a self-report survey instrument that assesses motivation along two dimensions, Effort and Importance. TTM was represented by a numerical score range for the subscales of Effort and Importance.
2. Effort: “The perceived degree of work or mental taxation put forth in completing the test” (SOS Test Manual, p. 5). Effort was quantified as the total of the ratings for the SOS items in the corresponding subscale, the range which is between 5 – 25.
3. Importance: “How important doing well on the test is to the student (the consequence

of the test for the student),” (SOS Test Manual, p. 5). Importance was quantified as the total of the ratings for the SOS items in the corresponding subscale, the range for which is 5 – 25.

4. Test type and stakes: in this study, “test type” referred to the distinction of a test according to the amount of consequence and/or reward for the test taker. This study used two such distinctions: “high-stakes,” which referred to a test having significant consequence and/or reward, and “low-stakes” which referred to a test lacking significant consequence and/or reward. The high-stakes test in this study was a college admission test such as the ACT or SAT, and the low-stakes test was the WV mandatory assessment.
5. College bound: in this study, the term “College bound” referred to a student who has taken the ACT/SAT or who plans to take it. “Non-college bound” referred to a student who has not taken the ACT/SAT and does not plan to take it.
6. Future plan: in this investigation, “future plan” referred to the students’ plan of being college-bound (participation in college admission testing) or non-college bound (non-participation in college admission testing).
7. Grade levels: in this investigation, “grade levels” referred to grades 9, 10, 11, and 12 in the sample.
8. Sex: in this investigation, “sex” referred to the high school boys and girls in the sample.

Methods

Sample

This study was conducted using a quantitative survey design with a target population of about 3000 students from five high schools in Harrison County, West Virginia. The approximate grade-level sample sizes needed from the population was derived using a sample

calculator set at a confidence level of 95%, plus or minus 5%. The necessary number of subjects was estimated using the calculator, and included the total number of students in each grade level and the associated p level (.05). Once the target sample was determined for each school, proportional sample sizes were then calculated for each grade level.

Variables

Test-taking motivation is the dependent variable and is operationalized as the Effort and Importance subscores measured by the Student Opinion Scale (SOS). These scores were causally related to the potential influence of four independent variables: test type, future plan, grade level, and sex. The influence of test type was the major independent variable, which was distinguished by the factors referred to as “low-stakes,” noted as the state-mandated achievement test, and “high-stakes,” represented by the college admission tests. An academic variable, future plan, was investigated as a potential influence on state testing motivation for the high school students, factored by those indicating college-bound and non-college bound. Additionally, demographic variables of grade level and sex were also considered as potential moderators of motivation on both test types. For the low-stakes test, grade levels will include 9 through 11, and for the high-stakes test, grade levels will include 9 through 12.

Data Collection

The SOS was administered by paper-and-pencil. The wording of the SOS was adapted in three ways to refer to the recently taken state test, the most recently taken college admission test, and for a planned admission test in the future. The statistical samples were randomly selected according to the desired fraction of the total population of each stratum within each school.

Delimitations

1. This study investigated the perspectives of the specific sample of high school students in Harrison County, West Virginia in the spring of 2017. The results of this study may be limited in generalizability due to the small sample size.
2. Assumption: the current study will consider admission testing as only a general indicator of a college-bound future plan. It has been found that all students who take or plan to take a college admissions test, about 33% of them do not expect to attend college (U.S. Dept. of Education, 2004).
3. Assumptions about the tests. This study assumed that the state assessment, regardless of test publisher or format, carried low stakes for students because it was mandatory, and there was no overt reward or consequence for performance level. Regarding the college admissions tests, this study assumed that high school students perceive no significant difference between the ACT and the SAT. Furthermore, because both admission tests are voluntary and because performance on either test could potentially help or hinder college plans, admission testing is considered to carry high stakes for students.

Summary

In the current era of educational testing and accountability, policy makers and administrators need to be confident in the data that drives their decisions. But in West Virginia, that data is obtained using a large-scale low-stakes test. Therefore, it is reasonable to consider whether or not students find the test important enough to try their best. This study attempted to assess this issue by comparing student motivation on the state test to motivation on a truly high-stakes test, the college admissions test. Questions regarding test-taking motivation have become increasingly prevalent since the early 1990s, and educational research has been gaining ground in

this area up to the present. However, there are few direct, large-scale comparisons between high- and low-stakes tests. Results from this study can contribute to this growing body of research by addressing these gaps.

CHAPTER TWO: REVIEW OF LITERATURE

Historical Context of K-12 School Testing

The following exposition will review strands of literature that demonstrate the importance of testing to various levels of U.S. education by outlining the presence of a testing culture, trends in federal government involvement, and recent state-level action. First, a discussion of the history of testing in U.S. schools will exhibit the dependence on assessments that has developed since colonial times and continues today. Then, a timeline of federal government involvement is divided into three distinct time periods during which federal pressure and accountability have increased and have progressively heralded additional testing. Prior to 1964, federal involvement took the form of legislation for various areas of education reform such as accessibility, opportunity, and curriculum. The period between 1964 and 1980 was marked by increases in federal demands for U.S. education systems, demands which most often stemmed from the need for accountability based on federal funding. From 1980 to the present, the level of federal pressure for school performance has continued to increase through the tactics of financial incentive and explicit legislation. The review is completed by a description of the most recent state-level actions regarding increases in federal pressure on testing. These topics support the relevance of the current study's overall purpose by demonstrating the complexity and magnitude of the reliance on educational testing in the United States.

The Testing Culture in the U.S.

History. The history of qualifying tests is a long and varied one, beginning with the famed Chinese civil servant examinations to the advancement of craftsman in the medieval craft guilds to the oral examinations in universities, popular from medieval times to the early 19th century (Hanson, 1993). America's history of dependence on testing in public schools began in

the mid-19th century when Horace Mann promoted the use of common written exams rather than the traditional model of university oral exams. This practice grew from Mann's vision of social advancement through schooling which, in turn, required some objective measure of the quality of learning and instruction and some way to compare schools and teachers. Large-scale examinations provided such measures since those early test results revealed education gaps, and as a result, further testing was developed to determine readiness for academic promotion. The success of Mann's model led to widespread use of written exams in the U.S., and his concepts formed the basis for the New York Regents Exams in 1865 (Gallagher, 2003).

Although the initial purpose of testing throughout history has been to sort or rank candidates according to prescribed qualifications, critics have expressed concerns that test preparation led to the tests becoming ends in themselves (Hanson, 1993). This same criticism arose early in the history of written academic tests because the focus of examination gained priority over subject matter and provided a basis for teacher and school evaluation which, in turn, led to "teaching to the test" (p. 199). However, these criticisms were confronted with practicality in U.S. history because there was a clear need for efficient ways to sort students effectively and objectively compared to the time necessary to conduct oral examinations and to the inconsistent nature of subjective evaluation.

Gallagher (2003) noted several factors that further encumbered the subjective assessment process. For example, the popularization of the elective curriculum increased the diversity of coursework and therefore increased the complexity of student assessment. Other factors such as immigration and child labor reform resulted in higher student populations and therefore increased the amount of work required to carry out assessments. For example, in 1845, Boston Public Schools instituted its written examination to test its 7000 students because this would

have been a daunting task using oral exams which were far more time-consuming and labor-intensive (Hanson, 1993). Although some feared that testing could shift the focus from educating to sorting and ranking, “the convenience of using objective tests outweighed this skepticism” (Gallagher, 2003, p. 85). Just as the Chinese civil servant examinations endured for a thousand years despite the same criticism (Hanson, 1993), the needs for examination in more modern times continued to trump the concerns for accurate evaluation.

In the early 1900s, the scope of large-scale testing increased with Edward Thorndike’s scientific approach of measuring human qualities, the development of intelligence tests, and the U.S. Army’s testing program for placement of soldiers during World War I (Gallagher, 2003). According to Rothman, the paper-and-pencil multiple-choice Army Alpha Test was objectively scoreable and recordable, and its format became the model for all standardized tests to follow (1995, in Gallagher). In the early to mid-1900s, several standardized tests emerged such as the Stanford Achievement Tests, the Scholastic Aptitude Test, the Iowa Test of Basic Skills, the Iowa Test of Educational Development, and the American College Testing Assessment. All of these tests could be administered and scored on large scales and easily provided a simple objective measure of knowledge and skill. Their value to education is evident because some of these tests and their revisions have endured late into the 20th century (Gallagher, 2003; Hanson, 1993), and some into the 21st century.

Modern testing industry. As schools and districts developed a dependence on large-scale testing to measure academic performance, the publishing industry created a niche of test developers to meet the demand. To execute yearly assessment, state departments of education have typically contracted test management companies, with contracts totaling about \$1.7 billion per year according to a 2012 (Chingos) estimate. This industry extends beyond test development

and publishing; for example, the membership of the Association of Test Publishers (ATP) includes dozens of companies involved in some or all areas of test development and testing support services such as delivery, scoring, and security (ATP, 2013). Like in many industries, some companies in this competitive field dominate, sometimes through diversification of purpose, if they can operate on large scales. These few companies win multi-year contracts valued in the hundreds of millions of dollars. In particular, Houghton-Mifflin, CTB McGraw-Hill, and Harcourt control most of the market, but others such as ETS have made significant gains. Hoff noted that although smaller companies sometimes win state contracts, the previously mentioned publishers have dominated the large-scale testing market in precollegiate schooling for the previous two decades (2003).

The field of test development has acquired a new form of non-corporate competition – consortia of states that are conforming to recent education reform in part by developing their own assessments. Several years of voluntary state-led efforts resulted in the development of standards aligned with demands of college and careers (Achieve, 2008). Leading the way in that reform was the National Governors Association Center for Best Practices (NGACBP) and the Council of Chief State School Officers (CCSSO) which published the Common Core State Standards (NGACBP & CCSSO, 2010). This reawakening of standards-based education reform has created the need for a new wave of Common Core assessments which have been developed by two consortia of states, Partnership for Assessment of Readiness for College and Careers (PARCC), and Smarter Balanced Assessment Consortium (SBAC). These groups will provide serious competition in the business of testing since development costs are covered by federal funding, and state-funded implementation costs are expected to be lower than current testing programs for many states (Doorey, 2013).

Trends in Federal Government Involvement

Before 1964: legislation for education reform. In U.S. history, the establishment of a structure of school governance occurred from local to state to federal jurisdictions. The period from Colonial times to the Civil War was marked by the passage of local education laws providing opportunity and the formation of governing bodies. During this time, colony and state laws required that communities provide schooling, and local boards of education formed when educational considerations such as funding and hiring began to overload local governments (Goldin, 1999).

Throughout the early era of American education, the action at all levels of government occurred as funding to increase accessibility and opportunity. However, the scope of government action expanded when states began forming departments of education in the 1800s, and following the Civil War, the U.S. Government formed a federal department of education to “collect information on schools and teaching that would help the States establish effective school systems” (U.S. Department of Education, 2012, para.4). By the late 1800s, the limitations of collecting and dispersing statistics about schools became evident regarding reform efforts, limitations which led to an early example of federal influence in higher education. In the late 1880s, the federal agency attempted to force improvement by ranking women’s colleges, and again in 1911, it attempted to rank a larger sample of colleges and universities. However, this later attempt was suppressed, due in part to the obvious complaint of negative publicity and to the fact that the rankings were based on a single criterion (Webster, 1984).

The 1900s began an era of interplay between state and federal governments regarding changes in accountability and increases in pressure on education; some actions at one level would influence actions at the other, and vice versa. The early 1900s was marked in part by

accountability through reporting of school test scores. For example, by the end of the 1930s, several states had testing programs in place, with many more to follow during the postwar years (Resnick, 1980). However, this supervision remained at the state level for several decades until new purposes for federal pressure emerged. One such purpose was motivated by the launch of Sputnik, after which the National Defense Education Act of 1958 designated federal funding to “strengthen the national defense” by focusing support in the curriculum areas of math and science in public schools and by boosting support for research and development in higher education (NDEA, 1958, p.1580). This legislation heralded further federal pressure on education, leading toward the creation of a national assessment.

1964-1980: Federal demands in public education. Both the level of government involvement and the focus of gathering public school data officially changed in the 1960s when data about “school buildings” and “how many years children stay in school” were deemed insufficient in the modern era (Keppel, in Vinovskis, 1998, p.5). This sentiment led to the development of the National Assessment of Educational Progress (NAEP) which was notable in two ways. Foremost, the NAEP was the first significant application of large-scale student assessments since the early intelligence and qualifying tests. In addition, the NAEP represented a continuation of federal involvement in curriculum which began with NDEA.

Considering its technical features, data, and political concerns, the NAEP was not a typical assessment. One major difference between the NAEP and previous assessments was the sampling method used. As a member of the development committee, prominent educator Ralph W. Tyler recommended using matrix sampling, which is a method to assess a large sample of students by having some students complete some parts of the whole test. Another major difference was the level of data reporting. Some educator alliances and even several state

government officials were strongly opposed to score reporting at school or state levels, so a compromise resulted in data being reported publicly for only four regional levels. A third major difference was that the NAEP had become a tool in the political arena. By the early 1970s, the federal government had acquired policy and financial control over the NAEP and began contracting the development and delivery to test development agencies (Vinovskis, 1998).

Although the federal government began imposing demands on education through the NAEP, its “definitive entry into public education” was the Elementary and Secondary Education Act (ESEA) of 1965 (Hanna, 2005, para. 10). This legislation continued previous reforms by further opening access to education and more importantly, to education funding. Its major provisions included funding for disadvantaged children, desegregated school districts, and non-public schools (Hanna, 2005; U.S. Department of Education, 2012). So, with this legislative tactic, the ESEA began an era of government involvement in education characterized by exchanging funding for compliance with reform goals.

The interplay between federal and state accountability continued after the federal government’s increased involvement during the 1950s and 1960s as many states began to mandate competency testing programs in regard to high school graduation and grade promotion. For this purpose, “the tests allowed the schools to show the tax-paying public that graduates had been prepared to function at a demonstrable level of competence” (Resnick, 1980, p. 9). Resnick further claimed that mandated minimum competency testing was, in part, a consequence of changes in budget management practices of the federal government in the 1960s. During this time, the U.S government began focusing heavily on objectives, analysis, and outcomes regarding Department of Defense budgeting. This focus on cost-benefit analysis spread in popularity through other government branches and then further to public and private agencies

throughout the U.S. State accountability legislation was an extension of this practice, and minimum competency testing was one example of that legislation (1980).

After 1980: increased federal pressure on school accountability. Federal education policy moved to the forefront with the establishment of the Department of Education as a Cabinet-level agency in 1980 and with the issue of the *A Nation at Risk* report in 1983. Although the state and federal interplay on education continued, the level of federal involvement in education can be characterized by tightening accountability and increasing pressure.

As states began developing assessments throughout the 1970s and 1980s, early criticisms about state-level data resurfaced, and NAEP results were found to lack the substance necessary for federal policymakers to carry out education reform. So, in later decades, the view of the 1960s was reversed, and state-level data reporting was deemed necessary for the NAEP (Vinovskis, 1998).

Regarding the state testing programs, the minimum competency form of accountability during the 1970s was short-lived, due to the increased federal influence in the early 1980s. *A Nation at Risk* claimed that these exams “fall short of what is needed, as the ‘minimum’ tends to become the ‘maximum,’ thus lowering educational standards for all” (U.S. Department of Education, 1983, p. 18). The influence of this federal report was enough to compel state education systems nationwide to abandon minimum competency testing in favor of the input-focused reform which centered on federal resource incentives such as increases in funding and staffing for schools that achieve higher levels of student proficiency (Kress, 2011). The resulting action from *A Nation at Risk* was a significant return to the financially motivated reform tactic initiated by ESEA in 1965. However, the emphasis on federal financial inputs led to

unsatisfactory results, so before the decade had passed, emphasis began to return to the outputs of education on the state level.

This standards-based reform focused on content and performance standards, along with the measurement and disclosure thereof. Some states even included such follow-up policies as recognition for success or accountability for failure in standards achievement (Kress, 2011). This combination of standards-based education and state-level accountability gained momentum in the late 1980s and 1990s, but the reform movement solidified even further as the state consequential accountability programs became a federal mandate with the 1994 reauthorization of ESEA known as the Improving America's Schools Act (IASA) (Kress). This Act combined the goals-and-outcome approach of the minimum competency movement with the financial incentives approach of the inputs-based movement by mandating that all states wishing to receive grants under the law adhere to certain accountability requirements. Such requirements include maintaining a measurable pattern of test score improvement and establishing contingencies for school districts that do not maintain that pattern. The details of these requirements are found in the following discussion on IASA.

Modern federal accountability in detail. Defined in Section 1111, IASA requirements included state content and performance standards, statewide assessments linked to the standards and deemed the primary determinant of yearly performance. This section also set forth a regimen of accountability beginning with “continuous and substantial yearly improvement” in state assessment scores, called adequate yearly progress (AYP) (Section 1111(b)(2)(B)(i)). The accountability program also mandated that performance results be reported annually upwards to the Secretary of Education with the result of local recognition or sanctions based on yearly progress. For exceeding AYP for three consecutive years, schools would be recognized as

“distinguished schools” (Section 1117(c)(2)(A)), and notably successful teachers in those schools may be recognized as “distinguished educators” (Section 1117(c)(3)(B)). But for failure to meet AYP for two consecutive years, schools or local agencies would be penalized with increasing severity (Section 1116). The first sanction level was for a school to be publicly identified as failing and then to develop and implement a plan of improvement. And for failing to meet AYP even after implementing a plan of improvement, the next level was to endure corrective action by the local agency. To be clear, the form of corrective action was up to the agency, because the law stated, “actions . . . which may include” followed by a list of suggestions such as the withholding of funds, the changing of school staff, decreased decision-making authority, and alternative governance arrangements (Section 1116(c)(5)(B)(i)). Similar to the consequences for schools, IASA set forth consequences for local agencies should a state agency determine district level failure in maintaining adequate progress. Local agencies were subject to the same identification, improvement plan, and corrective action pattern as were schools.

Enacted in 2002, No Child Left Behind (NCLB) reauthorized ESEA once again, strengthening the mandate in IASA regarding consequential accountability by increasing the level of sanction for poorly performing schools. Specifically, NCLB replaced the suggestive language in IASA regarding corrective action with more demanding language and added another level of consequence. The language of NCLB requires a local agency to “take at least one of the following corrective actions,” followed by a list of actions similar to those in IASA (Section 6313(b)(7)(C)(iv)). In addition to the school improvement and corrective action levels, NCLB provides for a more severe level of sanction in the form of restructuring, which involves significant changes in the decision-making structure, or governance. Again, the language of the

law is demanding in tone; in response to continued failure after corrective action, local agencies “shall implement” an alternative chosen from a specific list (Section 6316(b)(8)(B)). That list contains alternatives to current school and district governance, and this component has been a notorious consequence for consistently poor performance because it means that school leaders may have to turn over school operations to another entity such as a private management company or to the state education agency.

Regarding federal accountability, one may ask if it was necessary to increase the severity beyond the corrective action set forth in IASA? An analysis by Hanushek and Raymond (2004) of NAEP data comparing states with accountability to states with consequential accountability indicates “the force of accountability comes from attaching consequences to school performance” (p. 2). This analysis determined that consequential accountability made a positive impact on school performance, both overall and in some student subgroups. Specifically, their analysis determined that the amount of time that a state operated under such performance consequences was a significant indicator of NAEP improvements in math during the 1990s. Had this point of research been pursued earlier, perhaps NCLB would not have added a steeper penalty in terms of state assessments. However, within the variety of choices for school restructuring under NCLB, there appears to be a path of least resistance, a somewhat ambiguous option requiring “any other major restructuring of the school's governance arrangement that makes fundamental reforms . . . to improve student academic achievement” (Section 6316(b)(8)(B)(v)). Rather than choosing a state takeover of operations or replacing staff and administrators, schools may decide on some other form of alternative governance for a trial period. One commentator expressed that these “other” options chosen by schools are not as severe as the law intended and are merely repetitions of the severity of corrective action (Mead,

2007). Not surprisingly, the “any other” option has been the most popular choice among the states with schools in restructuring because this option holds the last opportunity for self-determination at the local level (Center on Education Policy, 2008). Whether the efficacy of consequential accountability has expired or schools are choosing soft restructuring options, the data on turnaround schools lacks promise (Hamilton, Heilig, & Pazez, 2013).

Federal pressure through competition. In 2009, the federal government augmented its consequential accountability reform with a return to the input-focused reform of the early 1980s by which major change would result from financial incentive. The American Recovery and Reinvestment Act authorized a competitive grant program, known as Race to the Top, for states “that are creating the conditions for education innovation and reform” (U.S. Department of Education, 2010b, p. 19496). The Assessment Program portion of this competition provided grants to two consortia of states totaling about \$360 million to develop assessments for states to comply with No Child Left Behind using the Common Core State Standards (U.S. Department of Education, 2010a)

Recent State-Level Action

State-level resistance. Recent trends in state education mandates center on NCLB waivers. Only three years after the enactment of NCLB, states began requesting alternatives to parts of the law. In January 2005, Virginia and Connecticut became the first states to request exemptions from NCLB requirements. Virginia’s request focused primarily on the provision of certain supplemental services, and Connecticut’s request sought exemption from expanding the scope of the state’s testing system (Hoff, 2005). The National Conference of State Legislatures (NCSL) issued a report questioning the constitutionality of a certain aspect of NCLB while encouraging the U.S. Department of Education to grant waivers provided by Section 9401 of the

law itself (NCSL, 2005). Requests similar to those of Virginia and Connecticut became increasingly common, although the following months were marked by conflict as procedural norms for the waiver process had yet to be established (Curriculum Review, 2005). Surprisingly, part of the conflict was open defiance from the state most closely associated with the passage of NCLB. Texas had overtly defied the NCLB requirement concerning the proportion of test exemptions due to learning disabilities. Connecticut's request for exemption from test expansion was rejected (Curriculum Review) as was California's request concerning testing of certain students with limited English language skills in reading and writing (Zehr, 2005). By accepting Virginia's request for autonomous decision making on certain school services, the U.S. Department of Education provided the impetus for several other states to submit proposals as well (Olson, 2005).

By late 2005, it was very clear that procedural norms were necessary as numerous questions arose concerning what constituted allowable alternatives to the law. During the next several years, as the U.S. Department of Education accepted and rejected more state requests, education officials worked to formulate specific waiver allowances and procedures. Meanwhile, Congress occasionally entertained the idea to revise the Elementary and Secondary Schools Act (ESEA) again, but it seems a revision may not be necessary. By 2011, the Department of Education had established ESEA flexibility "in exchange for rigorous and comprehensive State-developed plans" that remain loyal to ESEA core principles (U.S. Department of Education, 2013, p. 1). This flexibility waiver has a duration of two years, and the latest update allows for an extension of two additional years (2013).

District takeovers. In West Virginia, examples of the federal demand for corrective action may be found in several counties such as McDowell, Mingo, Grant, Hampshire, Lincoln,

and Preston. The state board of education has taken over the governance of these school districts for various reasons such as budget mismanagement and poor student achievement (West Virginia Education Association, 2009). Despite the strong language of the law, there is no federal mandate regarding the duration of alternative governance, nor is there provision for ineffective takeovers. An example of extended duration due to ineffectiveness is the oversight of McDowell County Schools, which lasted for over a decade until 2013 when the WV Board of Education returned control to local administration. Another example of ineffective takeover is the control of Mingo County Schools, where effects from the lack of local control have prompted a delegate from that county to sponsor a bill to limit the duration of a takeover by the state Board of Education to five years (West Virginia Legislature, 2014a). As of this writing, House Bill 4336 has passed and is under consideration in the Senate (WV Legislature, 2014b).

Increasing Year-End Test Score Performance

Strategies to Improve Test Scores

Gulek (2003) assembled from the literature a set of test-preparation strategies that promote sound pedagogical practices such as promoting time management and motivation, and reducing test anxiety. Many schools have held pep rallies to get children excited to take their state exams, but the literature on the effectiveness of this practice is lacking. However, there is no shortage of criticism for education in our testing culture. For example, Madaus and Russell (2010/2011) discussed multiple studies that demonstrate that the pressure to test is narrowing learning and teaching within subjects, across subjects, and across grade levels. Other effects include educational triage, or special attention paid to certain students to increase test results.

Incentives/Consequences for Students

Fryer (2010) describes a large-scale study which offered incentives for academic exercises but produced no significant result on state test scores. Regarding state tests in California, the practice of rewarding students for progress is becoming more popular despite the motivation theory criticism of extrinsic rewards undermining intrinsic motivation. The practice has extended to offering incentives just for participating in testing, with participation rates sharply increasing (Kuznia, 2012). In Chicago, Levitt et al. (2012) conducted an experiment that manipulated the motivation to perform well on a test by offering incentives for improving scores on a regularly given test. The participants of this study included nearly 7,000 students from over 30 schools in poorly performing school districts in the Chicago area, and the variables included incentive type, timing of the incentive offer, student grade level, subject area, and gender. Levitt et al. offered two financial incentives, \$20 and \$10, and one non-financial reward, a trophy. These rewards were offered at three points in time, a delay of one month, immediately following the test, and before the test. The offer before the test served to apply behavioral economics to the testing environment by framing the incentive as a loss – the students who improved their test scores were allowed to keep the incentive. As one would expect, the delayed rewards yielded no improvements. Of the immediate rewards, only the \$20 incentive led to significant improvement at 0.12 to 0.2 standard deviations. Regarding the incentive-as-a-loss, students in the elementary grades improved performance by 0.18 to 0.25 for a trophy, and students in high school improved performance by 0.12 to 0.13 for financial rewards. The authors concluded that educators may derive some benefit of cost-effectiveness using the trophy reward framed as loss when assessing students in grades two to five, but this aspect of behavioral economics may not be a reliable motivating tactic for older students (2012).

Incentives/Consequences for Teachers

A current trend in education is to connect teacher compensation to student performance, and as Fryer (2011) reported, the trend is occurring in many countries and across the U.S. This practice is opposed by teacher unions (Fryer) but is supported by the federal government, as expressed in NCLB, a state may “recognize and provide financial reward” to high-performing teachers (Section 1117(b)(3)). An enduring example of a merit pay system is the education system of North Carolina where incentives are paid at the group level to all teachers in each school that performs at level (Ahn & Vigdor, 2011).

Student Motivation for and Interest in Test Achievement

Historical Approaches to Motivation

Mathematical. The application of probability to human behavior began in 1654 with an investigation of gambling by the mathematicians Blaise Pascal and Pierre de Fermat (Brandstätter, Gigerenzer, & Hertwig, 2006). Their idea of expected value was defined as a sum of probabilistic monetary values for each possible outcome, leading to a gambler’s choice of the option having the highest expected value. This explanation of rational choice was actually an early theory of reasoning, and the first significant adaptation occurred in 1738 when Daniel Bernoulli replaced the objective monetary values with subjective utilities in terms of a logarithmic function. Due to certain trends of philosophical and political thought during the late 19th century and early 20th century, theorists of human thinking ignored probability and Bernoulli’s expected utility (2006). But beginning in the 1930s, probabilistic expectations, along with expectancies and values, reappeared in the mathematical and quasi-mathematical models used in the motivation field theory of Lewin (1935), the drive theory of Hull (1943), the resultant valence theory of Lewin, Dembo, Festinger, and Sears (1944), the axiomatic approach to

economic behavior of von Neumann and Morganstern (1947), the social learning theory of Rotter (1954), the achievement motivation theory of Atkinson (1957), and the expectancy theory of Vroom (1964). Selected examples of probability are found in the motivation models of Atkinson and Vroom. Atkinson's model asserts that the subjective probability for success, P_s , together with the achievement need, M_s , and the incentive value of success, I_s , are factors of the tendency, T_s , to approach a goal in his equation (Graham & Weiner, 1996):

$$T_s = M_s \times P_s \times I_s$$

Vroom's (1964) model is similar:

$$MF = E \times I \times V$$

In this formulation, the expectancy, E , is the subjective probability that one's effort will lead to the desired level of performance; the instrumentality, I , is the subjective probability that performance level corresponds to magnitude of the reward; and the valence, V , is the subjective value of the reward (Scholl, 2002). Although the construct test-taking motivation would not surface for decades after Atkinson and Vroom, one can clearly see the connection between the past and present concepts, namely, expected probability for successful performance and value of the reward.

Psychological. Two major psychological perspectives in motivational theories have been the belief that humans are driven to satisfy needs and the question of how much control humans have over the factors that influence their behavior (Gollwitzer & Oettingen, 2002). In the research involving needs, there has been variety in the number and arrangement of needs. Two models that exemplify the variety in number are Clark Hull's drive theory based on a single need and Henry Murray's system of 20 psychogenic needs (2002). Hull conceptualized need as a "general, content-free drive" (p. 10110) that energizes behavior toward some physiological

deficit (Graham & Weiner, 1996). In contrast, Murray distinguished between 20 needs that motivate and direct behavior; his catalog included the needs for achievement, autonomy, and understanding (Schultz & Schultz, 2009). Regarding the question of control, early theories usually considered the internal and external influencers of human behavior to be compulsive in nature, and more modern theories acknowledge various levels of personal input into those factors that influence behavior (2001).

Achievement motivation. According to Graham and Weiner (1996), the modern field of motivational psychology includes a variety of concentrations but has become largely centered on achievement. This convergence on the topic of achievement began when the team of McClelland, Atkinson, Clark, and Lowell (1953) performed the first systematic study of achievement motivation (Weiner, 2013). McClelland and his team focused on Henry Murray's (1938) need for achievement, and using Murray's Thematic Apperception Test (TAT), they found that the expression of the achievement need is able to be influenced by cues in the task environment. Specifically, the experimenters gave instructions that reflected various levels of achievement and found that responses on the TAT generally matched the instructions. This empirical work was closely followed by Atkinson's (1957) construction of a theory of achievement motivation which applied constructs similar to those in Lewin's field theory to individual differences in achievement strivings (Weiner, 2013).

The following overview of various achievement motivation theories begins with expectancy-value theories and proceeds according to central constructs in the separate areas of expectancy and value. The discussion concludes with a description of various other background factors of achievement motivation.

Expectancy-Value Theories of Achievement Motivation

Atkinson. Contrary to Lewin's (1935) field theory, John Atkinson's (1957) achievement motivation theory focused on individual differences in motivation. However, Atkinson used constructs nearly identical to Lewin's and applied these to achievement. Atkinson's construct called tendency represented behavior attempting an achievement goal, and in his model, tendency is directly related to each of motive, probability, and incentive:

$$T_s = M_s \times P_s \times I_s$$

Motive, or need for achievement, was assumed to be "an enduring disposition to strive for success" (Graham & Weiner, 1996, p. 70). The expectancy variable was represented by the construct he called probability of success and was often defined operationally in terms of the difficulty of a task. The value variable was represented by his construct called incentive and referred to an individual's value of success. An important feature of Atkinson's model was his formulation that the incentive value of a task is negatively related to its probability of success because a low probability of success should elicit a greater feeling of accomplishment and therefore, a greater incentive value. For example, in the classroom, Atkinson's model would predict that a student would find more value in accomplishing advanced tasks compared to simple tasks because accomplishing advanced tasks would place the student in an elite group possessing greater skills than the majority of classmates.

Eccles et al. Research during the decades following Atkinson's theory has led to an elaborate expectancy-value model for achievement motivation (Eccles et al., 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). In this model, Eccles and her colleagues have differentiated several factors into their constituent factors and have delineated chains of causality. The result is a complex system comprised of five layers of causality, beginning with

cultural milieu and stable child characteristics and ending with achievement-related behaviors. The system contains at least 25 separate factor dependencies, so for the sake of this discussion, only the relationships involving expectation of success, subjective task value, and achievement-related behavior will be considered.

Achievement-related behavior, comprised of task choice, persistence, and performance, is assumed to be directly influenced by expectancy and value. This model distinguishes subjective task value into the components attainment value, defined as personal importance of a task; intrinsic value, defined as enjoyment and/or interest; utility value, defined as the relation of a task to an individual's goals; and cost, defined as the negative aspects of task engagement, which include effort (Eccles & Wigfield, 2002). In a causality test of their model, Eccles et al. (1983) demonstrated that expectancy was caused by self-concept of ability, a self-assessment of competence in performing a given task, which was caused by perceptions of cost and task difficulty.

This study included various perceptions and experiences regarding high school mathematics for males and females over two years. Among the many results obtained by Eccles et al. (1983), two are directly relevant to the current strand of discussion. First, their causal analysis suggested that a student's perception of the value of an activity is the significant determinant of the decision to engage in the activity. In the high school setting, this result would apply to those activities for which the student has a choice such as future course taking but would not apply to those activities such as mandatory assignments and standardized tests. The other specific result from the Eccles et al. study is that once a student is involved in an activity, one's self-concept of ability is a significant determinant of performance. In the high school

setting, it stands to reason that this result would apply to all activities, both voluntary and mandatory.

Models of Feather and Heckhausen. Other modern work in expectancy-value theory of achievement motivation since Atkinson (1964) focuses on many ideas similar to those of Atkinson and/or Eccles et al. (1983) but place different emphasis on certain areas (Eccles & Wigfield, 2002). For example, Feather (1988) emphasized value by broadening this concept to include an individual's set of values and arguing that these values influence the valence of goals (Eccles & Wigfield, 2002). Heckhausen (1991) made a similar contribution by emphasizing expectancy. His model distinguished four types of expectancies, each defined as a subjective probability focused on outcomes (Eccles & Wigfield, 2002). The first expectancy is a situation-outcome connection for which a student would consider the chance of reaching a specific outcome by doing nothing. During a semester of coursework, a student may consider whether or not skipping an assignment would lead to a lesser course grade or even whether putting forth no effort on a statewide assessment would lead to any less desirable outcomes in future schooling. Heckhausen's (1991) action-outcome expectancy is similar to Vroom's (1964) expectancy variable; this is a student's expected chance of reaching a specific outcome through certain actions. An example of this occurs when students consider whether a certain amount of effort on a statewide assessment would lead to a certain performance level. Heckhausen (1991) combined the previous two expectancies into a third called action-by-situation-outcome, which is a student's perceived chance that factors within the situation would influence one's action-outcome expectancy. The final expectancy in Heckhausen's resembles Vroom's instrumentality variable; this is a student's expected chance that a particular outcome will lead to a certain consequence. An example of this expectancy occurs when students take statewide assessments

and may ask “What reward will I get if I score well?” and “What happens to me if I score badly?” It would seem that the administration at school, district, and state levels have considerable control over this expectancy because policies at these levels can guarantee rewards and consequences for any performance level on the high-stakes assessments.

Achievement Constructs Focused On Expectancy

Expectancy for success is the perceived probability that certain behaviors will lead to a desired goal, and an individual’s expectancies are beliefs about one’s performance on given tasks (Eccles & Wigfield, 2002). This perception depends on an individual’s needs and perceived level of control. In particular are one’s self-judgment beliefs, and these beliefs are framed in a variety of ways such as self-worth, outcome expectation and self-efficacy, helplessness, and locus of control.

Self-worth. The approach in the self-worth theory of Covington and Beery (1976) proposed that individuals strive to maintain a sense of self-ability because society places high value on the ability to achieve. The central construct in this theory is ability self-perception, and the central belief is that the primary motivator of behavior is the need for individuals to perceive themselves as competent (Graham & Weiner, 1996). This concept manifests in the school testing arena as a student’s need to see oneself as competent, and the intensity of this need would motivate a corresponding level of effort on an annual assessment.

Self-efficacy. Other constructs that contribute to expectancy are Bandura’s concepts of outcome expectation (1997) and self-efficacy (1977). Outcome expectations are beliefs that “certain behaviors will lead to certain outcomes” (Eccles & Wigfield, 2002, p. 111). Self-efficacy refers to individuals’ beliefs about their potential to perform well on a given task (Graham & Weiner, 1996). This construct has been quite valuable in the field of achievement

motivation for several reasons. First, self-efficacy is a robust variable representing self-confidence, expressed along the dimensions of strength, generality, and level (Eccles & Wigfield, 2002). Furthermore, an individual's belief in one's potential to perform a task addresses the traditional theme of control over factors that influence behavior (Gollwitzer & Oettingen, 2001). And specifically, in terms of motivation, Bandura (1989) claimed that "People's self-efficacy beliefs determine their level of motivation, as reflected in how much effort they will exert . . ." (p. 1176). And finally, according to Graham and Weiner (1996), self-efficacy has more consistently predicted behavior and change in behavior than other similar expectancy variables.

Helplessness. The negative version of both outcome expectation and self-efficacy can be expressed in terms of Seligman's (1975) concept of helplessness. This learned phenomenon results from an individual's negative internalization of past failures, the effects of which can include loss of self-esteem, lowered expectancy for future effort and outcomes of similar events and even of events in other contexts (Graham & Weiner, 1996). Of all connections to the school testing arena, the example of helplessness affecting test effort may be the most prevalent and therefore the most obvious explanation for low test-taking motivation. In this example, a student who has scored poorly on standardized tests in the past may give up on trying to succeed to avoid trying and failing. A student in this position would decide that it is just as easy to fail without effort as it is to fail with effort. This perceived chance of achieving the same outcome by doing nothing is an example of Heckhausen's (1991) situation-outcome expectancy.

Locus of control. According to Eccles and Wigfield (2002), the locus of control theories developed by Crandall et al. in the 1950s and Rotter in the 1960s began a line of research in the belief that expectancy is directly related to an individual's feeling of being in control or out of

control, of one's success and failure. Although early versions of the theory focused only on internal and external control over success, more recent work in this area has developed the concepts much further. For example, Connell (1985) asserted that unknown control can be a factor by demonstrating that not knowing the causes of success or failure can undermine motivation to perform given tasks (Eccles & Wigfield, 2002). Later, Connell and Wellborn (1991) integrated the traditional themes of needs and control with a model that proposes a set of needs – competence, autonomy, and relatedness – that mediate the effect of control beliefs on motivation. Their model states that control beliefs are linked to the competence need and that fulfillment of the set of needs will motivate behavior (Eccles & Wigfield, 2002). Fulfilling a need for competence in a school setting is an application of self-worth theory for students with an internal locus of academic control. In this example, locus of control theory would assert that fulfilling the need for academic competence should motivate effort on a standardized test.

Attribution. Similar to the concept of locus of control is the concept of causal attribution in the attribution theory of Weiner et al. (1971). Attribution theory incorporates aspects from other theories of motivation and behavior, but central to this model is an individual's perception of the causes of past successes and failures and the relationship this perception has on present and future tasks. Both locus and control of causes are aspects of the attribution model, but are considered separately, along with the stability of a cause (Graham & Weiner, 1996).

For example, a student sitting for a test may consider each of these three aspects regarding past success or failure on similar tests. Regarding locus of cause for success or failure on previous assessments, a student may reflect on whether one's actions led to the outcome or if test performance was caused by some outside influence. A student attributing success or failure to an internal cause may also consider controllability of the cause by asking "What did I do that

led to my test score?” Similarly, a student attributing success or failure to an external cause may also ask “What did someone else do or what else happened to lead to my test score?” And finally, a student may consider the stability of the cause of test performance by reflecting on study habits, for example.

Achievement Constructs Focused on Value

Value is the value an individual finds in a behavior and/or its consequences. Like expectancy, value is a perception, and it depends on an individual’s needs. Specifically, theoretical models have focused on the location of the source of a goal’s incentive value, intrinsic or extrinsic.

Intrinsic incentive. Regarding a given activity, an individual who is intrinsically motivated derives the incentive value from within; in other words, the activity satisfies an interest and/or provides some enjoyment (Eccles & Wigfield, 2002). The concept of interest has been developed in recent decades in terms of situational interest, initiated by specific features of a given task, and individual interest, a “relatively stable evaluative orientation” toward a given task (Eccles & Wigfield, 2002, p. 114). Schiefele (1999) has distinguished between two types of attractive features, or valences, which comprise individual interest. Feeling-related valences refer to a task’s associated feelings, and value-related valences refer to a task’s significance to the individual. Nicholls (1984) focused on goal orientations in his formulation of achievement motivation. Concerning the intrinsic incentive, individuals may derive a task’s value in terms of competence relative to that of others, which is called ego-involvement, or they may determine value relative to previous levels of mastery, which is known as task-involvement. According to Deci and Ryan’s (1985) self-determination theory, a given task has intrinsic value if the incentive satisfies the psychological needs for competence and autonomy.

There has also been some focus on individual differences in intrinsic motivation as an individual's inclination to rely on intrinsic incentive, which can be linked to need for achievement (Eccles & Wigfield, 2002). The notion of intrinsic incentive rests squarely on the central topic of the current study as a major factor in test-taking motivation. As students approach their annual assessments, all educational stakes and accountability through every level of personnel and administration must focus on the belief that students are intrinsically motivated to follow directions, do what is expected of them, put forth sincere effort on the test, and want to score well. Many schools have accepted the fault in this logic and have tried to counter the potential lack of intrinsic incentive by offering external incentives such as rewards for high achievement test scores or even consequences such as mandatory enrollment in future remedial courses.

Extrinsic incentive. In contrast, an extrinsically motivated individual receives the incentive from without in the form of a reward or other benefit (Eccles & Wigfield, 2002). Graham and Weiner (1996) noted the well-documented phenomenon that an external reward can decrease the intrinsic incentive of a task. Deci and Ryan's (1985) self-determination theory accounts for this phenomenon with the idea that an individual may internalize an external reward through interpretations of competence or autonomy. These theorists classified external rewards according to function, specifying that rewards are controlling if the effect is a perceived influence on behavior, and that rewards are informational if the effect is feedback about competence (Graham & Weiner, 1996). Deci and Ryan (1985) argued that rewards internalized as controlling will not satisfy the need for autonomy and would therefore undermine a task's intrinsic incentive value. The notion of extrinsic incentive complements the concept of intrinsic incentive in the current study. Examples in the school testing environment can be found in every

high school in the U.S., but specifically, extrinsic incentives for test performance would appear as rewards offered by school administrators or as consequences such as remediation. However, the assertion of Deci and Ryan (1985) that the students' perceptions of the test rewards may undermine the value of the rewards. For example, students may believe that avoiding remediation is a reward in itself but may see it as a controlling type of reward designed as bait for the students to score well on the test. In this case, which is not far-fetched given the rebellious reputation of adolescents throughout history, the students in question may decide not to put forth a good effort.

Background Factors of Motivation

Socializers. Several results have been reported regarding various background factors that influence aspects of motivation. One major study is that of Eccles et al. (1983) which offered an elaborate expectancy-value model of achievement motivation by investigating several connections between beliefs of students, teachers, and parents regarding ability, difficulty, and importance of math. The sample was comprised of 668 students in grades five through twelve, along with their parents and teachers, and the data was collected in the forms of student records, classroom observations, and questionnaires for all participants. Part of the analysis compared the influence of parents and teachers on students' math self-concepts in a variety of ways. For example, they found that the effect of parents, and especially mothers, on achievement-related beliefs was stronger than that of teachers. Furthermore, the influence of parents was found to be a result of their role as "direct socializers of achievement beliefs and attitudes" rather than their "power as role models" (p. 137). To accentuate this result, Eccles et al. noted that, in the test of the correlations between parents' self-concept variables and student responses and student math performance, none of the correlations was significant.

Age. Graham and Weiner (1996) report results from several studies that show younger children are less vulnerable to the negative effects of “maladaptive motivational patterns” (p. 80) such as low expectancy of success, causal attribution to uncontrollable factors, and concern with preserving self-worth. Also, these studies show that younger children have high self-concept of ability, have not yet learned behaviors of helplessness, and have not yet developed a perception that effort is inversely related to ability.

Gender. In a 1984 paper, Eccles, Adler, and Meece elaborated on the Eccles et al. (1983) survey data to focus on differences between males and females regarding achievement choice. One of their goals was to delve deeper than a common report on gender difference data. Their data suggests that the sex differences in course taking decisions are influenced in different ways. Analysis showed that past academic performance was a strong factor influencing course taking choices for both males ($p < .0004$) and females ($p < .0005$), but for males this was the only significant factor. However, they found that subjective task value is a strong and independent factor ($p < .004$) for females’ course taking choices.

Gender differences in self-concept of ability, intrinsic value, and academic performance have also been shown to exist. Yoon, Eccles, and Wigfield (1996) used data collected for the Michigan Study of Adolescent Life Transitions (MSALT), which conducted questionnaires with 826 students at four points in time over two school years. They found that boys benefitted academically from the promotion of positive and unrealistic self-concept of ability, and that girls benefitted from promotion of realistic self-concept of ability. They also concluded that intrinsic value had consistent positive influence on performance.

Race and ethnicity. Similar to the results of Eccles, Adler, and Meece (1984) who strove to determine underlying influences on gender differences, there are factors shown to influence race and ethnic differences in school motivation. One example is that types of parental involvement are linked differently to measures of school motivation for different race and ethnic subgroups (Fan, Williams, & Wolters, 2012). This result was found using base data from the Educational Longitudinal Study of 2002 and survey responses from a national sample of 12,721 tenth graders and their parents. The surveys involved five aspects of parental involvement, five indicators of academic motivation, and four ethnic groups. Relevant to the current study are the motivational variables of self-efficacy in English and in math, along with intrinsic motivation in those subjects. Of the 100 possible correlations, dozens of effects were noted, and results were mixed. One example is the effect of parents' aspiration for their children's postsecondary education which had overall strong positive association with the motivation aspects across the ethnic groups. Another consistent result is the negative influence of parent-school communication regarding problems on all motivation variables, where significant. In contrast, the effect of parental advising showed wildly mixed results regarding self-efficacy and intrinsic motivation in both math and English across the ethnic groups.

Expectancy-Value Theory of Achievement Applied to Test-Taking

The expectancy-value framework is well-suited to the explanation and assessment of test-taking motivation due to the commonly measured constructs of effort and importance (e.g., Sundre, 1999; Wise & DeMars, 2003). In particular, testing effort and importance of the test may be considered identical to the cost and attainment value aspects of subjective task value, according to the Eccles et al. (1983) model. From this perspective, test-taking motivation is based in the value portion of expectancy-value theory, with influence from the expectancy factor

due to the causality of self-schemata and goals on subjective task value (Eccles & Wigfield, 2002). One example that provides evidence that the model is appropriate is the distinction of test-takers into three effort categories related to goals orientation and personality types (Barry, Horst, Finney, Brown, & Kopp, 2010).

Effort. In a study grounded in the expectancy-value theory of motivation, Abdelfattah (2010) investigated the association of test-taking motivation and performance on low-stakes examinations. With a sample of 797 Arabic ninth graders randomly chosen in 11 schools, the author measured the “effort scores” students put forth on math or science exams for which they knew had no consequence for their school grades. Participants responded to the Student Opinion Scale, translated into Arabic, which assesses levels of student effort and test importance during a given test session. Although the means for this data indicated that importance of students doing well on the test was higher than their effort in taking it, students who reported greater importance on the test tended to exert greater effort. A comparison of the subscores showed the correlation to be high, $r = .612$ for math and $r = .572$ for science ($p < .01$). Students whose global motivation scores were greater than one standard deviation from the mean were classified as low- or high-effort, accordingly. The mean test scores for both subject areas were significantly higher for the high-effort group, and effect size was found to be high, $\eta p^2 = .272$ for the math test and $\eta p^2 = .207$ for the science. These results indicated that overall, low effort significantly decreased test performance.

Computer-based tests provide the ability to measure item response time, which is interpreted as response time effort (RTE), and since the 1990s, rapid-guessing behavior has been studied as a detrimental factor because the resulting response correctness is essentially due to chance (Wise, Bhola, & Yang, 2006). Using analysis of RTE, Wise and DeMars (2008) have

pinpointed the item location at which a test taker abandons test effort. And as a measure of test-taking effort, item response time has been shown to agree with a self-report measure, the Student Opinion Scale (Sundre & Moore, 2002), for approximately 66% of a particular sample (Swerdzewski, Harmes, & Finney, 2011).

Importance/value: “high-” vs. “low-stakes.” To describe test importance, the terms “high-stakes” and “low-stakes” are commonly used ambiguously. To whom are the tests important? One answer is that tests are important for those whose goals are served. Who pays the price for “high-stakes”? These are serious questions in a testing culture with an increasing level of federal pressure and related testing mandates in various states. To clarify the ambiguity of “stakes,” several authors have acknowledged that the modern level of accountability requires tests with little or no consequence for the test-takers but with great consequence to schools (e.g., Swerdzewski, Harmes, & Finney, 2011; Wise, Kingsbury, Thomason, & Kong, 2004). Wise et al. noted further that “tests that are low-stakes to students are particularly vulnerable to the effects of lack of student effort” (2004, p. 9). These conclusions render low-stakes test data quite valuable because from the test-taker’s perspective, even so-called “high-stakes” tests may actually be non-consequential.

Wise and DeMars (2003) suggested raising the stakes by imposing consequences for test performance, a suggestion supported by Thelk, Sundre, Horst, and Finney who found high test-taking motivation on a “very high-stakes assessment” (2009, p. 145). Similar to raising test stakes is the practice of increasing the level of control in the testing environment, as in Barry and Finney’s 2009 study of proctored college examinations. In a post-secondary gender study of test-taking motivation under low-stakes conditions, DeMars, Bashkov, and Socha (2013) found that women put forth greater effort, and showed evidence that certain personality traits were

likely factors. In this study, test-taking motivation was measured using the SOS and RTE with agreement between the measurements. For example, there was a difference in both measures of motivation that indicated men tended to expend lower effort than women at low levels of motivation. Furthermore, on the SOS, there was greater variation in the males' effort than that for females (variance ratio = 1.34).

Threats to validity for achieving optimal student scores. Concerning test-takers in an era of accountability, Swerdzewski et al. (2011) asked, "If they are not exerting effort, what does that mean for the test results used to inform policy decisions?" (p. 163). An earlier statement by Wise and DeMars (2003) could provide an answer, "If low motivation leads to reduced test scores, it also reduces score validity in the most basic sense" (p. 10). Wise, et al. (2004) pinpointed a greater problem and suggested a level of remedy, "Given that low student motivation can pose a threat to the validity of inferences made concerning assessment test results, it is important to assess the degree of effort students expend toward their test" (p.3).

Sundre and Wise (2003) proposed a procedure called motivation filtering to remove the adverse effect of low effort on a set of test scores, which in turn could reduce the distortions in the assessment of proficiency levels. They further suggested that it may also be useful to retain only the high-motivation data. Wise, Bhola, and Yang (2006) identified the adverse effect as construct-irrelevant variance (CIV) and determined that effort-monitoring computer-based tests reduced this variance. They warned that non-consequential tests are sensitive to effort-related CIV, naming as examples the K-12 state tests, the NAEP, international tests of student achievement, and pilot tests for new assessments.

Measuring Student Motivation for Test Taking

Student Motivation Surveys

Within two decades following Atkinson's (1957) landmark work on the need for achievement, there appeared a variety of surveys and tests to measure the achievement motive. Some examples are the French Test of Insight (1958), the Mukherjee Sentence Completion Test (1965), the McReynolds and Guevara Success-Seeking and Avoidance of Failure Tests (1967), the Costello Task-Orientation and Success-Orientation Tests (1967), the Mehrabian Test of Resultant Achievement Motivation (1968), the Achievement Orientation Scale (1971), and the Ray Achievement Motivation Scale (1975) (Ray, 1979; Veroff, McClelland, & Marquis, 1971). These and more recent surveys vary widely in number of items and number of subscales. Some examples of this variety are the Revised Achievement Motives Scale with only 10 items and two scales (Lang, 2006), the Ray Achievement Motivation Scale with 14 items and only one scale (Ray, 1979), and the Achievement Motivation Inventory with 170 items and 17 scales (Schuler, Thornton, Frintrup, & Mueller-Hanson, 2004). Like any theorist, the authors of these instruments have pinned their focus onto a conceptual framework they believed to be important. To simplify the lot, one may assume that the factors in these surveys are variations on Atkinson's (1957) dichotomy of the achievement motive into an approach tendency, called hope of success, and an avoidance tendency, called fear of failure (Lang, 2006). Versions of these two concepts can be applied to the school testing environment in the form of expectancy and value.

Assessments of Test-Taking Motivation

An early assessment related to test-taking motivation is found in a 1984 attitude survey by Karmos and Karmos regarding the perceived importance of standardized tests, a conclusion from which is that "students tended to react negatively to standardized achievement tests"

(Stefanou & Parkes, 2003, p. 153). Notable results are that a staggering 47% of students in grades 6-9 thought that taking these tests was a “waste of time,” and 21% reported low effort on achievement tests (Kiplinger and Linn, 1993). However, following Karmos and Karmos, the constructs of test value and test-taking effort remained unapproached for several years. The Test Attitude Survey (Arvey, Strickland, Drauden, & Martin, 1990) provided another relevant measure of testing motivation but was developed external to the educational arena (Schmit & Ryan, 1992).

The NAEP studies. It appears that former American Federation of Teachers president Albert Shanker is responsible for the surge of measuring educational test-taking motivation that began in the 1990s when he called for study into effort on the NAEP in his New York Times column:

If students know that what they do on a test doesn't matter, they may decide it's not worth their while to put forth any effort. . . . NAEP is an important source of information about what U.S. students know and can do, so we ought to clear up this question about its validity (July 29, 1990).

Shanker's recommendation was to conduct an experimental study manipulating the stakes of NAEP by offering a prize for good performance and counting test scores toward student grades.

Throughout the decade following Shanker's public call, a flurry of investigations ensued. Early measures of low-stakes test effort and value are found in NAEP mathematics field test data (Educational Testing Service, 1991). Relevant to the current study are the field test questions about perceived effort and importance. When asked, “How hard did you try on this test?” 28% of 8th graders and 51% of 12th graders responded with “Somewhat hard” or “Not at all hard” (O'Neil, Sugrue, Abedi, Baker, & Golan, 1997, p. 2). In response to the question, “How

important was it for you to do well on this test?” 36% of 8th graders and 62% of 12th graders chose “Somewhat important” or “Not very important” (p. 2). These early results are important for three main reasons; first, the results justified Shanker’s publicized questions about the value of tests to students and the amount of test-taking effort being put forth. Second, although the investigations began in force with the NAEP, researchers later sought to generalize the results to other low-stakes tests such as statewide standardized achievement tests. Finally, the results gave credibility to a line of research which began with an attitude survey regarding standardized tests in general (Karmos & Karmos, 1984), which could then be expanded to involve surveys given in conjunction with tests to determine the effect on performance (e.g. O’Neil, et al., 1997) and later could be developed into a published product to measure test-taking motivation specifically (Sundre & Moore, 2002). The following review will discuss several self-report instruments designed to assess test-taking motivation.

In 1992, O’Neil, Sugrue, Abedi, Baker, and Golan conducted a series of experimental studies involving test-taking motivation and NAEP performance. The portion of these investigations most relevant to the current study is their working model for test-taking effort, which modified existing models of expectancy-value achievement motivation to involve gender, grade level, and ethnicity variables along with experimental manipulation of incentives for test performance (O’Neil, et al., 1997). The referent test was a released portion of the 1990 NAEP math test, and the self-assessment questionnaire for 8th grade participants in this study consisted of 25 items and addressed two metacognitive variables along with self-reported effort and worry. The survey for 12th graders consisted of 35 items and addressed four metacognitive variables along with self-reported effort and worry. O’Neil et al. found several gender and ethnic differences at each grade level and found correlations among the variables along with test

performance. A relevant result is that females in both grades reported greater effort than males. Furthermore, for the 8th grade sample, the treatment of financial incentive and an increase in perceived effort were accompanied by an increase in performance. However, in the 12th grade sample, the treatment of financial incentive was accompanied by an increase in other metacognitive variables but not a significant change in test performance (1997).

Beyond the NAEP. Two surveys of Paris, Lawton, and Turner (1992) departed from the NAEP investigations but enhanced the results of Karmos and Karmos (1984) by determining a significant age difference. Paris et al. (1992) conducted one study using a survey of attitudes on standardized tests in four states with nearly 1,000 students in grades 2-11 and found that older students reported less motivation to do well on tests than younger students. The other was a survey of 250 students who took a state-mandated test in Michigan in grades four, seven, and 10 with a result that older students, compared to younger students, reported placing less value on the test and “filling in the bubbles without thinking” (Kiplinger and Linn, 1993, p. 7).

Brown and Walberg (1993) did not use a survey but still contributed to the literature of the time with a study that manipulated the intrinsic motivational condition of a test. Using an encouraging script of instructions in the experimental group, Brown and Walberg found a highly significant effect ($F = 10.59, p < .01$) with an increase in test scores of .303 standard deviations. This study marks a turning point in the history of measuring test-taking motivation, beyond which measurement became more sophisticated.

Test-taking motivation questionnaires. Wolf and Smith (1995) conducted a study to investigate the relationships between consequence, motivation, and test performance. However, the significance of this study lies in the method they used to measure test-taking motivation specifically. To achieve this, they developed the Student Motivation Questionnaire with eight

Likert-type items that focused on importance of test, effort on test, and value of test results. They implemented this survey to a pool of psychology majors in conjunction with a test given in an undergraduate psychology course, and their operational definition of consequence centered on whether or not the test counted toward the course grade rather than on other external motivators. The motivation scales showed reliabilities of $\alpha = .84$ for the consequential condition and $\alpha = .89$ for the nonconsequential condition, and further analysis supports a unidimensional motivation scale. The main difference between this survey and that of O'Neil et al. (1992) is that Wolf and Smith (1995) focused more specifically on motivation with a single factor survey whereas O'Neil et al. analyzed the interaction of effort with several other factors in the same survey

As this line of research developed during the late 1990s, focus remained on the questions of test validity and test score interpretation. In 1997, O'Neil et al. issued their final report on the NAEP validity studies begun in 1992 and stated the implication, "The 8th-grade findings indicate that, indeed, we may be underestimating the achievement of students when we use scores on 'low-stakes' tests as indicators of achievement" (p. xii). Around the same time, the work of Donna Sundre (1995, 1996, 1997, & 1999) replicated the work of Wolf and Smith (1995) and refined the survey instrument by explicitly identifying importance and effort as significant aspects of examinee motivation (1997). Sundre (1997) found that importance accounted for 49.9% of variance, and effort accounted for 19.9% of the variance in her 1995 data. The result of this series of studies is the Student Opinion Scale (SOS), with 10 items surveying the two factors, Importance and Effort. In 2002, Sundre reported that the SOS has been administered to over 15,000 college students and its scales consistently have had reliabilities in the .80s. The simplicity, reliability, and relevance of the SOS to the central topic of the current study make it the most natural choice for the data gathering instrument in the current project.

Despite the success and ease of the SOS, it is not the only way to assess test-taking motivation. For example, Eklöf (2006) studied a sample of 343 Swedish eighth graders who took the 2003 Trends in Mathematics and Science Study (TIMSS), a low-stakes test for the participants. To measure motivation, she developed her own Test-Taking Motivation Questionnaire (TTMQ) grounded in the expectancy-value model of achievement motivation (Eccles & Wigfield, 2002). The survey consisted of 24 items, including one open-ended and three categorical items, with ordinal items offering four or five response alternatives. Factor analysis revealed three scales consistent with expectancy-value theory, labelled General Attitudes, Performance Expectancy, and Test-Taking Motivation (TTM). Although attitudes and expectancies fit into the theoretical model, most relevant to the current study are the items weighting on the TTM factor, those regarding students' perceived importance of and effort on the TIMSS. Eklöf (2006) found that about 17% of the variance was attributable to the TTM factor, and this scale had an acceptable reliability at $\alpha = .84$.

Response Time Issues

According to Luecht and Sireci (2011), computer-based testing (CBT) has been a viable alternative to traditional paper-and-pencil testing for four decades. During this time there is one particular adaptation that brings CBT directly into the focus of the current study – the ability of the testing software to record a tester's response time. Wise and Kong (2005) capitalized on this ability of the software to develop a new measure of a tester's effort. Response Time Effort (RTE) provides information item-by-item without using a self-report instrument to judge one's effort after the test is finished. The software can record the response time as the difference in seconds between an item's appearance on screen and the student's input of a solution. These response times can be analyzed item-wise or in aggregate and can then be compared to some

agreed upon threshold of appropriate time to read and respond to a particular item. Wise and Kong (2005) identified rapid-guessing behavior as an indicator of low testing effort and tested RTE compared to the Effort subscale of the SOS. Their analysis of results from a sample of 472 college freshmen showed that RTE agreed with self-reported effort on the SOS; for example, an SOS mean of 12.50 was found for examinees with RTE scores less than .80, but an SOS mean of 16.38 was associated with RTE scores greater than .90.

This line of research has continued to develop. For example, programming a warning system into a CBT can improve testing effort. Wise, Bhola, and Yang (2006) demonstrated this improvement from a sample of 318 college sophomores taking an effort-monitoring CBT. Specifically, the mean RTE for the control group (no warning for frequent rapid guesses) was .82, and the mean RTE for the warning group was .88. And to help prevent testers from giving up on exerting effort on a CBT, another statistical analysis has been found to approximate the item at which a tester begins to engage in rapid-guessing behavior. Wise and DeMars (2010) have attempted to apply the technique called rolling person fit to identify the point during the test when the percent of correct answers begins to resemble chance-based guessing. The results were not solid, but the attempt serves as an example of the potential of CBT.

Motivation Filtering

Sundre and Wise (2003) proposed a technique called motivation filtering which is aided by RTE but is possible using any measure of test-taking motivation. This technique is the process by which test scores are reported and interpreted only for examinees identified as having expended adequate effort. The purpose is simple: to increase validity by aggregating only those scores which represent effort on the test. In the landmark study of this process, Sundre and Wise administered a battery of tests, including the SOS, to two samples of nearly 800 college students

and compared these scores to those students' SAT scores. To begin the filtering process, the SOS scores were grouped into six categories of effort, and validity statistics were computed for successive removal of low-scoring groups. The results of this process clearly demonstrate its value. As successive groups were removed, there were corresponding increases in the correlations between SAT scores and performance on the assessment. The variance shared by SAT and test performance showed the result as $r^2 = 20\%$ for the total sample, and $r^2 = 41\%$ for the most strictly filtered sample.

Summary

The historical dependence on assessments that has developed in the U.S. occurred alongside marked increases in federal government involvement in education. Especially during the second half of the 20th Century, federal pressure and accountability have increased and have progressively heralded additional educational testing. To accommodate the increasing dependence on testing and the changing curriculum paradigms, the modern testing industry has matured into a billion-dollar-a-year business with multi-state contracts worth hundreds of millions of dollars. The states must compete in several ways to fulfill their educational duties: for federal funding, for test publisher contracts, and to meet performance accountability requirements.

With this hype and pressure on administrators and educators, what pressure is on the students who are the mediators in this massive process? Students as test-takers are the connection between the efforts of schools and the accountability requirements of state and federal governments. Motivation theories have been developed and tested to explain what is likely going on in the minds of the students. Surveys have been developed and tested to identify the answers. The literature clearly expresses the concerns of educators that something is

missing: a guarantee of student effort on assessments that carry such high stakes for the professionals. And the lack of such a guarantee leads to a questionable level of confidence in test score interpretation. To help contribute to this confidence, a direct comparison of current test-taking motivation on large-scale low- and high-stakes tests is needed because the literature is incomplete regarding this type of study. The current study will attempt to address this gap in West Virginia by comparing motivation on the low-stakes state-mandated achievement test, the Smarter Balanced Assessment, and the high-stakes college admission tests, the ACT and SAT.

CHAPTER THREE: METHODS

Purpose

The main purpose of this study was to determine what differences exist in test-taking motivation among high school students in both low- and high-stakes testing situations; in particular, for those completing the annual WV state-mandated achievement test and for those completing a college admission test. A closely related purpose was to determine what differences exist in state testing motivation between college bound and non-college bound students. An additional purpose was to determine any differential effects of these outcomes across several grade levels, and finally, it was of interest to know if the issues were moderated differently for boys and girls.

Study Design

The design was empirical, using a quantitative survey of self-perception administered to a random sample of high school students in one Appalachian school district. The descriptive results served to gauge test-taking motivation, an aspect of education that has received little attention in the state included in the study. Further statistical analysis compared the dependent variable to each of the independent variables to carry out the central and related purposes of this investigation.

Variables

The dependent variable was a measure of test-taking motivation with two subscales, *Effort* and *Importance*, each keyed to a numerical rating agreement subscale from 1 - 5. The major independent variable was test type with the factors of *low-stakes* and *high-stakes*. Additional independent variables included future college plan, (factored by participant indication

of college-bound or non-college bound), student grade levels and sex. These variables were considered as potential moderators of test-taking motivation on both test types.

Limitations

1. A central limitation is the use of self-report data which depends on the assumption that respondents were willing and motivated to complete the survey honestly regarding self-perception of motivation on certain tests.
2. Because the participants of the study were high school students who received the survey at the schools attended by the students, permissions from county and school administrators, along with parental consent, had to be obtained. With several levels of permissions and consents, it was possible that participation was limited.

Population and Sampling

The target population of this study consisted of 2,969 students from five high schools in Harrison County in north central West Virginia (Harrison County Schools Central Office, 2016). The county has approximately 69,000 residents and median household income of about \$43,000 (U.S. Census Bureau, 2013).

Survey Population

The feeder area for each high school has distinct geographic and population differences including rural, suburban, urban, along with several small towns ranging from affluent to poor. The five schools with enrollments by grade level are listed in Table 1.

Table 1.

Harrison County high school enrollments by grade level (2016-2017)

	9	10	11	12	Total
Bridgeport HS	203	216	178	157	754
Lincoln HS	151	160	128	123	562
Robert C. Byrd HS	210	193	195	152	750
South Harrison HS	87	96	102	83	368
Liberty HS	159	140	122	114	535
Totals	810	805	725	629	2969

Source: Harrison County Schools Central Office

Sample Sizes

The necessary sample size from the entire population of 2,969 students was determined using an online sample calculator set at a confidence level of 95%, plus or minus 5% (Creative Research Systems, 2012). The calculator yielded a sample size of 340 students, and this result was doubled and rounded up to 750 students to help offset a potentially high rate of non-returns. The sampling scheme was streamlined by sampling classrooms randomly, assuming an average class size of 25 students. The ratio of each school’s population to the total population was multiplied by 750 to yield the sample of students, and then divided by 25 and rounded up to yield the number of classrooms required per school. Because of rounding up, 32 classrooms were randomly selected. Finally, the ratio of each grade level population within each school was multiplied by the sample per school to yield the sample of classrooms per grade level.

For grades 9-11, the classrooms are homerooms. But for the 12th grade students, the use of homerooms was not always feasible because many seniors leave school for work or for college classes. Therefore, when available, the 12th grade participants were surveyed in their English 12 or Civics classes.

Procedures

To provide support for conducting this study, approval was granted by the candidate's doctoral committee and the Marshall University IRB, the district superintendent, and school administrators (see Appendices A – C). The 2016-17 enrollment of each approved school formed the total population for this study, and because most participants were minors, and parental consent was attained as well. To achieve this level of permission, consent forms were distributed to all students to take home. All students who returned affirmative consents and who agreed to participate received the survey.

On Site

Contact was made with each school principal to invite participation in this study. And for those administrators who agreed to participate, requests for parental consent were delivered to the schools for distribution to all included students. Meeting times were scheduled in each school to manage survey access for those students with affirmative parental consent.

Survey Delivery

The Student Opinion Scale was delivered as a paper-and-pencil survey to each student who returned an affirmative parental consent. The researcher handled the distribution and collection of the surveys in person.

Data Preparation

Regarding Research Question One, levels of TTM were ascertained by compiling the SOS subscores of Effort and Importance, and then scores from schools in each homogeneous demographic group were compared during the analysis. State-test TTM consisted of the separate Effort and Importance subscores for the entire homogeneous sample group, whereas admissions

test TTM consisted of the separate Effort and Importance subscores for the students in the group who had taken or planned to take those tests.

To address Research Question Two, each homogeneous sample group was split into two disjoint subsets according to the indication of a college-bound future plan on the survey.

Regarding Research Question Four: in WV, the state achievement test is not administered to high school seniors; therefore, the 12th grade responses were combined with the current 11th grade sample because this data represents reflections of their motivations levels as 11th graders.

Data Management

The SOS was administered only to those students under the age of 18 who have returned affirmative parental consents and have confirmed student assent to participate, along with those adult students who were issued informed consent statements. The importance of this exclusivity was stressed to all volunteers to adhere to the ethical protocols of consent. The data were input into an SPSS file. Protocols were set up within the file to reverse score the negatively worded items, according to Sundre (2007) and to sum responses to yield subscores for Effort and Importance.

Although the survey responses were anonymous regarding individual students, the data contained school identifiers. Therefore, in the data file, the school names were changed to bogus names. The data file was maintained securely on the researcher's personal home computer and backed up in cloud storage using Dropbox. All files were maintained as such until deemed no longer necessary, under the advice of the doctoral committee Chairman.

Instrument and Data Collection

The Student Opinion Scale (SOS)

The SOS is a self-report survey grounded in the Expectancy-Value model of achievement motivation (Eccles, et al., 1983). This model asserts that the constructs most immediately responsible for achievement-related behavior are the Expectancy for one's success at a task and the subjective Value of that task. In first developing the SOS based on a predecessor (Wolf & Smith, 1995) and further through her own revisions, author Donna Sundre has found that the Expectancy and Value constructs from the achievement motivation model manifest as Effort and Importance within the testing environment.

The SOS consists of 10 Likert-type items measuring Effort and Importance with a 5-point self-rating scale as follows: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree. The data obtained from the items was summed according to the relevant subscale as follows: items 1, 3, 4, 5, and 8 factor on the Importance scale, and items 2, 6, 7, 9, and 10 factor on the Effort scale. Each participant will score from 5 to 25 for each subscale, and upon recommendation from the author in the SOS Test Manual (2007), the two scores will be considered separately in this study, and negatively worded items, numbers 3, 4, 7, and 9 will be reversed prior to scoring.

Technical aspects. One aspect of the SOS bears directly on the focus of the current study – that of effect sizes of the difference in subscale scores for different test stakes. The effect sizes of the differences were calculated using a sample of seniors at a four-year college who took a high-stakes exam in their course of study and a sample of sophomores at a four-year college who took a “required battery of general education assessments in a low-stakes context” (SOS Test Manual, p. 6). The reported values of Cohen's d are 1.30 for the Effort subscale and 2.30 for

Importance, providing strong evidence that examinees report higher test-taking motivation in high-stakes settings.

Additionally, reliability estimates are reported in the SOS Test Manual for survey administration under both low- and high-stakes testing conditions at a four-year college and under high-stakes conditions for a community college sample. For the low-stakes settings with freshmen and sophomores at a four-year college, Sundre (2007) reported Cronbach alphas in the low .80s for the Importance subscale and in the mid-.80s for the Effort subscale. In high-stakes settings with seniors at a four-year college, Sundre noted low variability in the Importance subscale, which contributed to the low reliability values of .46 and .59. In contrast, the Effort subscale for these samples showed acceptable reliability values of .83 and .76. For the high-stakes settings for graduating students in the community college samples, Cronbach's alpha values were in the upper .80s for both Importance and Effort subscales. Reliability of scale was estimated for the current sample using Chronbach's alpha. The analysis was obtained by the intercorrelations of all items (effort and importance) for both the college tests and the state exam. An overall alpha of .823 (n, 139) was obtained for the college tests and an alpha of .836 (n,156) was obtained for the state test. These overall values indicate good internal consistency of the scale for the samples.

To validate the instrument, Sundre (2007) conducted the three stages of Benson's (1998) program of validation. First, in the substantive stage, Sundre established the specific connection to Expectancy-Value Theory, noting the limitations. For example, the Importance factor relates to the subjective task value component of the theory but does not further differentiate between types of task value. Therefore, this subscale indicates the student's general concept of the value of the test. Furthermore, the Importance factor is positively correlated to the Effort factor, but

the relationship is not causal. Although the Effort factor does indicate a level of expected success, it does not indicate specific causes for expectancies or for the amount of effort and persistence a student spends on a test.

The second stage addresses internal validity which was accomplished by a factor analysis of the original eight-item SOS, per the Motivation Scale (Wolf & Smith, 1995). In contrast to the single-factor structure of the predecessor, confirmatory analyses of the SOS supported a two-factor structure, prompting adjustments of the items and leading to the present form (Sundre & Finney, 2002; Thelk, 2004; Thelk, 2006).

The final stage is external validity, which Sundre supported by citing Wise (2006) who demonstrated SOS subscale correlation to certain external constructs. As would be expected from two separate measures of effort, Wise found moderate correlation between SOS Effort and response time effort ($r = .433$) and found a lesser correlation between Importance and response time effort ($r = .293$). The same study also compared the SOS scores with the Information Literacy Test (ILT) and the SAT total. Also, correlations were moderate between Effort and the ILT score ($r = .460$). The final comparison reported was a low correlation between Importance and ILT score ($r = .258$). Based on the evidence provided in the SOS Test Manual, the Effort subscale appears to have a higher level of external validity than does the Importance scale.

Survey Instrument

The SOS data collection instrument adapted for the current study is a two-part survey. Part I is a brief section, labelled “Student Information,” which asked about gender, grade level, and school. Part II is the test-taking motivation survey, the Student Opinion Scale (SOS), which is a pre-existing instrument that consists of 10 Likert-type items assessing Effort and Importance with the 5-point self-rating scale previously described.

To achieve the goals of this study, the SOS was presented in three different reference contexts: Survey #1, labelled “Main Survey,” asked participants to reflect upon the state test; “Survey #2,” asked participants to reflect upon a previous college admission test; and “Survey #3,” which asked participants to reflect upon a future college admission test. Instructions at each survey directed participants about the expectations of the survey. Non-college bound participants were instructed to complete only the “Main Survey” and stop. College-bound participants were instructed to complete the “Main Survey” and go on to complete “Survey #2” if they had taken an admission test or “Survey #3” if they had not yet but planned to complete an admission test.

Data Analysis

In raw form, the data consisted of all survey responses and were stored in the researcher’s home computer and backed up online using the researcher’s Dropbox account. Using SPSS Version 22 the matrix for each respondent consisted of grade level, sex, school, state Effort, state Importance, college plan, admission Effort, and admission Importance. For high schoolers indicating a non-college bound plan, the admission Effort and Importance cells were void and therefore ignored in the analysis. The grade level, gender, Importance, and Effort data represented the variables laid out in the research questions and were analyzed per each research question.

Research Questions

Question One. What are the levels of test-taking motivation (TTM) reported by high school students who complete the WV mandatory assessment and ACT or SAT college admission tests?

Question Two. What differences exist in TTM reported by high school students who complete the WV mandatory assessment and the ACT or SAT college admission tests?

Because the Effort subscores for state test completers are independent of the Effort subscores for admissions test takers, analysis consisted of an independent samples t test. When significance was found, an effect size measure was calculated to determine the proportion of variance in Effort scores attributable to each factor of the test type variable. Identical analysis was conducted regarding the Importance subscores reported by state test completers and admissions test takers.

Question Three. What differences exist in TTM between college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?

The mean sub-scale scores for Effort subscores reported by college-bound state test completers was compared to the mean of Effort subscores reported by non-college bound state test completers using an independent samples t test. When a significant difference was found, the effect size was calculated to determine the proportion of variance in state Effort subscores attributable to each factor of the future plan variable. Identical analysis was conducted regarding the Importance subscores for college-bound and non-college bound state test completers.

Question Four. What are the differences in TTM reported by college-bound high school students who complete the low-stakes WV mandatory assessment and by the same students who take the high-stakes ACT or SAT college admission tests?

Because the Effort levels reported by college-bound state test completers and those reported by college-bound admissions test takers are derived from the same set of participants, the analysis will consist of a paired t test. If a significant difference is found, the effect size will be calculated to determine the proportion of variance in college-bound Effort levels attributable

to each factor of the test type variable. Identical analysis will be conducted regarding the Importance subscores for college-bound state and admissions test takers.

Question Five. What grade level differences exist in TTM reported by students who complete the low-stakes WV mandatory assessment?

Question Six. What grade level differences exist in TTM reported by college-bound high school students who take the high-stakes ACT or SAT college admission tests?

Questions Five and Six address the comparison of Effort and Importance scores across grade levels and were analyzed by an analysis of variance (ANOVA) model to compare the means of the groupings. When evidence was found for a significant difference, post hoc analysis was conducted to determine its effect size.

Question Seven. What differences exist in TTM reported by girls and boys who complete the low-stakes WV mandatory assessment?

Question Eight. What differences exist in TTM between those reported by girls and boys who take the high-stakes ACT or SAT college admission tests?

Because the Effort subscores for girls are independent of those for boys, analysis consisted of an independent samples t test. Identical analysis will be conducted regarding the Importance subscores reported by girls and boys.

Summary

The main purpose of this study was to determine what differences in test-taking motivation exist among students in both low- and high-stakes testing conditions. The low-stakes scenario for all students in this study was the state-mandated achievement test, and the high-stakes scenario chosen for high school students in this study was a college admission test.

Within the high school segment, it was also reasonable to compare the test-taking motivation levels among the college-bound and non-college bound students.

Data were collected using a well-established survey designed for the purpose of measuring self-perception of test-taking motivation, and this survey was administered in Spring 2017. The format was paper-and-pencil, and the analyses consisted simply of comparisons of aggregated means and analyses of variance.

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

Introduction

Chapter Four presents the results of the data analysis and findings related to each of the research questions that follow. This chapter also provides a brief review of the study's data collection tool and related demographics, its population and sample, and its data analysis techniques.

Purpose

The central purpose of this study was to determine what, if any, differences exist in test-taking motivation among high school students for completing a high-stakes college admission test and the low-stakes annual WV state-mandated achievement test. More specifically, the assessment sought to extract potential differences in perception regarding the Effort expended for each of the measures as well as to yield an indication of the Importance of these tests to the individual. In effect, could these measures be perceived as having “high stakes” or “low stakes” for these participants? It was also determined if these outcomes were moderated by several demographic factors, including gender, grade level, and college plans.

Population and Sample

The sample for the study was composed of 161 high school students from a single Appalachian district. The sample included 75 females, 79 males, and 7 other participants who left the gender item blank. The grade level composition was as follows: 52 ninth grade, 45 tenth grade, seven eleventh grade, and 54 twelfth grade. The district includes five high schools, with the following sample participation: 51 from Central High, 4 from County High, 20 from Mountaintop High, 41 from River City High, 44 from Valley High, and 1 missing value.

Sampling results were disproportional for college plans with 140 for college-bound compared to 17 for non-college bound.

Data Analysis and Results

Research Question 1.

What is the level of Test-Taking Motivation (TTM) reported by students who complete the low-stakes WV mandatory assessment and high-stakes ACT or SAT college admission tests? Conceptually, TTM is the dependent variable in the study, and operationally, Effort and Importance are the related observable measures of TTM.

The Student Opinion Scale (SOS) measured two distinct subscales, Effort and Importance, each represented by five of ten items on the SOS. Subscores for these constructs were the sums of participant responses per the numerical scaling for its items (1 – 5), thus ranging from 5 to 25. Subscores with missing values are not included in the analysis. Table 2 summarizes the descriptive data for these variables.

Table 2

Descriptives for Level of Importance and Effort on State and College Admission Tests

	N	Minimum	Maximum	Mean	Std. Deviation
Admission test Effort	140	7	25	21.46	3.253
Admission test Importance	139	5	25	21.96	3.421
State test Effort	156	6	25	17.46	3.911
State test Importance	157	5	25	17.06	4.338
Valid N (listwise)	137				

The differing N-values for each factor of the dependent variable resulted from the inclusion of participants who answered all relevant questions on the survey. Results for the mandatory state achievement test yielded a mean score of 17.46 for Effort and a mean score of

17.06 for Importance. Regarding the college admissions test, mean scores of 21.46 and 21.96, respectively, were obtained for Effort and Importance. Variability scores were lower for Importance and Effort in the college admissions sample (mean = 3.34) compared to the state sample (mean = 4.13). The greatest variability score ($\sigma = 4.34$) overall occurred for Importance by those in the state sample. Overall, these data do indicate a trend: lower scores for Effort and Importance occurred for the state level exam compared to higher scores for Effort and Importance for the admissions test.

Table 3

SOS Item Statistics for State Tests^a

SOS Item	N	Mean	Std. Deviation
state1 Importance of doing well	157	3.59	1.230
state2 Engaged in good effort	157	3.65	1.103
state3r Curious about my score	157	3.23	1.260
state4r Concerned about score	157	3.50	1.264
state5 Important test to me	157	3.13	1.229
state6 Best effort [gave]	157	3.62	1.216
state7r Not work as hard as can	157	2.43	1.172
state8 Want to know how well	157	3.62	1.147
state9r Not give my full attention	156	3.35	1.328
state10 Able to complete test	157	4.41	.848
Valid N (listwise)	156		

a. r = Reverse Polarity Items

Table 3 summarizes individual item data for the State survey. Overall these data are similar with the exception of the mean for item # 7 (2.43), indicating respondents reported they did not work as hard as they could have on the test. Item 10, with a mean score of 4.41, shows that subjects were able to complete the test. Overall, these mean scores corroborate the mean score for State data noted in Table 2, which shows the lower results for State tests compared to College tests. Standard deviations varied considerably indicating a relatively high spread of

scores within a 5-point range. The spread could have been caused by several factors including outliers in the database and the fact that completing the survey did not hold any particular consequence for respondents. Additionally, effort and motivation differences may have existed among those in the various grade levels. Table 4 summarizes the individual item data for college admissions tests.

Table 4

SOS Item Statistics for College Tests^a

SOS Item	N	Mean	Std. Dev.
adm1 Importance of doing well	143	4.71	.627
adm2 Engaged in good effort	143	4.63	.670
adm3r Curious about my score	143	3.76	1.236
adm4r Concerned about score	142	4.36	1.074
adm5 Important test to me	143	4.60	.827
adm6 Best effort [gave]	143	4.67	.712
adm7r Not work as hard as can	143	3.62	1.412
adm8 Want to know how well	142	4.51	.800
adm9r Not give my full attention	142	4.19	1.193
adm10 Able to complete test	142	4.36	.832
Valid N (listwise)	141		

a. r = Reverse Polarity Item

Mean scores for these items clustered closely with the exception of items 3r (*Curious about score*) and Item 7r (*Not work as hard*). The latter was surprising given that the college test has significance for admission, potential scholarships and course placements. In addition, it was somewhat surprising that subjects scored relatively low (Mean, 3.76) regarding being curious about their scores given the presumed importance. These data are consistent with the overall mean scores for college tests noted in Table 2 when compared to the state test for Importance and Effort. There was lesser variability associated with the college results as the means clustered for the greater part. Yet some variability occurred for several of the items noted in the table.

Most likely, these were associated with outliers in the database and perhaps differences among those at the grade levels. For example, 9th graders may not have been concerned about college tests and scores at this point in time and varied in regard to the amount of attention and energy given.

Research Question 2.

For clarity, the wording of this research question in the discussion has been altered to distinguish between the results regarding differences in Effort and Importance. The data in Table 5 are summary of the data (averaged mean scores) for this analysis.

Effort – What differences exist in Effort reported by high school students who complete the low-stakes WV mandatory assessment and high-stakes college admission tests? This is a direct comparison of Effort level between the two test types.

A paired samples t-test for comparing the average mean scores for college and state tests yielded a significant t-score for Effort, favoring the college tests at the 95% confidence level of the difference ($p < .001$). Calculating Cohen's d using an online effect size calculator, the effect size for this result was $d = 1.08$ (www.uccs.edu, 2000). Taken together, these data indicate that respondents reported giving significantly greater effort when completing college admissions tests.

Importance – What differences exist in the levels of Importance reported by students who complete the low-stakes WV mandatory assessment and high-stakes college admission tests? This is a direct comparison of Importance level between the two test types.

A paired samples t-test comparing the mean difference for college admission tests and the state test means was significant at the 95% confidence level ($p < .001$). Calculating Cohen's d , the effect size for this result was $d = 1.22$. These results showed statistical differences with a

high level of magnitude, indicating that subjects reportedly attached greater importance to their performance on the college tests compared to the state tests.

Research Question 3.

For clarity, the wording of this research question has been altered to distinguish between the results regarding Effort and Importance. Table 5 contains the data for this analysis.

Table 5.

Descriptives for Future Plan, Effort, and Importance for State and College Tests

Test—E&I	Future Plan	N	Mean	SD	SEM
State—Effort	College-bound	139	17.51	3.929	.333
	Non-college bound	17	17.18	3.861	.936
State—Imp.	College-bound	140	17.20	4.339	.367
	Non-college bound	17	16.41	4.403	1.068
Adm—Effort	College-bound	140	21.39	3.253	.275
	Non-college bound	17	X	X	X
Adm—Imp.	College-bound	139	21.94	3.421	.290
	Non college bound	17	X	X	X

There are no calculated values for non-college bound respondents because these students did not take or do not plan to take a college admission test.

Effort – What differences exist in the levels of Effort among college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?

Concerning the Effort given on the state test, responses of college bound participants (N = 139) had a mean of 17.49, compared to a mean of 17.18 for non-college bound participants (N = 17). Using a t-test for equality of means, the difference was not significant at the 95% confidence interval of the difference. This result was the same whether or not equal variances were assumed. Overall, scores were not distinguished by college or non-college plans.

However, this outcome was likely affected by the disproportion of college bound to non-college bound respondents. Notably, scores for the college tests were greater than those for the state test.

Importance – What differences exist in the levels of Importance among college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?

Concerning Importance of the state test, college bound participants (N = 140) reported a mean of 17.14, compared to a mean of 16.41 for non-college bound participants (N = 17). These data were analyzed by a t-test for equality of means, and no significant difference was found in this relationship. This result was the same whether or not equal variances were assumed. Overall, respondents reportedly gave greater effort for their performance on the college admission tests compared to the state test; however, these results are interpreted cautiously given the disproportion in resulting sample sizes, 140 college bound to 17 non-college bound. Again, and notwithstanding the disproportions in the samples, college plans were not a significant factor in the rankings.

Research Question 4.

Effort – What is the difference in levels of Effort reported by college-bound students who complete the low-stakes WV mandatory assessment and by the same students who take the high-stakes ACT or SAT college admission tests?

College bound students (N = 140) obtained a mean of 17.51 on the SOS for state test Effort, compared to a mean of 21.39 for the college admissions test SOS. These data were analyzed with a paired samples t-test, which was significant at the 95% confidence interval of the difference ($p < .001$). Calculating Cohen's d , the effect size for this result is $d = 1.07$. Again, effort scores for college tests were significantly greater than those of the state test.

Importance – What is the difference in levels of Importance reported by college-bound students who complete the low-stakes WV mandatory assessment and by the same students who take the high-stakes ACT or SAT college admission tests? These data are displayed in Table 5.

The level of Importance on the SOS for the state test showed a mean of 17.20, compared to a mean of 21.96 for the college admissions test SOS. Conducting a paired samples t-test, significance occurred at the 95% confidence interval of the difference ($p < .001$). Calculating Cohen's d , the effect size for this result was $d = 1.21$. Again, these results indicate that greater importance is attached to the college tests.

Overall these data show that the levels of Effort and Importance are greater among respondents where college admissions tests were concerned, compared to the mandatory state level test. Additionally, the tests of significance indicate that these results are beyond chance, and that the magnitude of these differences is high, considering the effect size results.

Research Question 5.

What grade level differences exist among the levels of Effort and Importance reported by students who complete the low-stakes WV mandatory assessment?

As planned, the 11th grade results were combined with the 12th grade results, and this combination proved to be necessary because the 11th grade response was low, with only seven participants. Table 6 summarizes the descriptive data for these variables.

Table 6.

Descriptive Data for Grade Levels, Type of Tests, Effort, and Importance

Grade	Measure	Adm Effort	Adm Imp	State Effort	State Imp
9	Mean	22.06	21.62	18.31	17.50
	Std. Deviation	2.714	3.379	3.15	3.98
10	Mean	21.54	21.68	17.95	18.00
	Std. Deviation	3.822	4.122	3.97	4.28
11/12	Mean	20.87	22.50	16.35	15.64
	Std. Deviation	3.169	2.811	4.25	4.51
Total	Mean	21.46	21.96	17.46	17.06
	Std. Deviation	3.253	3.421	3.911	4.34

Mean scores for Effort on the state test ranged from 16.35 to 18.31 and from 15.64 to 18.00 for Importance. These data were analyzed by a one-way analysis of variance. Results showed significant differences for effort between 9th and 12th graders ($p < .023$) and between 10th and 12th graders ($p < .05$). In both cases, the trend was for scores to decrease slightly up the grade levels as the lowest scores, on average, occurred for 12th graders and the highest for 9th graders.

Research Question 6.

What grade level differences exist in the levels of Effort and in the levels of Importance reported by college-bound high school students who take the high-stakes ACT or SAT college admission tests?

Notably in Table 2, the mean scores for college tests for both Effort and Importance are very similar and averaged 21.46 and 21.96 respectively. The data in Table 6 indicated a decrease in SOS Effort scores upward through the grades; however, a one-way analysis of variance for these data showed no significant differences across the two outcomes, as expected. In short, respondents equitably reported high levels of Effort and Importance for these tests.

Research Question 7.

What differences exist in the levels of Effort and Importance reported by girls and boys who complete the low-stakes WV mandatory assessment? Table 7 summarizes the data for the Effort variable.

Table 7.

Descriptive Data for Sex by Effort and Importance on State Test

Sex	Measure	State Effort	State Imp
Female	Mean	18.00	18.01
	N	73	74
	Std. Deviation	3.648	4.100
Male	Mean	16.78	16.13
	N	77	77
	Std. Deviation	4.109	4.414
Total	Mean	17.37	17.05
	N	150	151
	Std. Deviation	3.926	4.352

Effort and Importance data were analyzed by an independent samples t-test and significance was found for gender, favoring girls ($p < .039$). However, an effect size measure of $d = .028$ indicates a very low magnitude of significance. In short, there is mathematical significance, but it is of little practical importance. Similarly, significance was found for Importance, again favoring girls ($p < .006$). A measured effect size of .05 likewise indicates a very low level of magnitude. In short, while the mean scores were slightly greater for girls, these differences were likely a chance factor in regard to effort and importance on the low-stakes state test.

Research Question 8.

What differences exist in the levels of Effort and Importance reported by girls and boys who take the high-stakes ACT or SAT college admission tests? Table 8 contains the descriptive data for the Importance outcome.

Table 8.

Descriptive Data for Sex by Effort and Importance on College Tests

Sex	Measure	Adm Effort	Adm Imp
Female	Mean	21.41	22.10
	N	70	70
	Std. Deviation	3.7	3.6
Male	Mean	21.50	22.08
	N	64	63
	Std. Deviation	2.82	2.9
Total	Mean	21.46	22.09
	N	134	133
	Std. Deviation	3.27	3.2

These data were further analyzed by an independent samples t-test and significance was not found for sex. Both girls and boys reported similar mean scores for college tests. In short, gender did not distinguish one's perceptions of the Effort given and Importance attached to the high-stakes college admission tests. However, as noted for the overall results, these groupings again confirm that higher levels of Effort and Importance self-perceptions are associated with college admissions testing.

Summary

Overall, the data indicate the following kinds of results and patterns. Aggregated mean scores for Effort and Importance were greater for the college admissions sample compared to the state exam sample. This difference was supported statistically, and the assumption is that the

latter gave lesser effort and attached lesser importance to the low-stakes state exams. Although there is a level of consequence to schools and school systems for not meeting statewide compliance policies, state exams apparently do not have much at stake for the students. There is no reward or punishment per se to the individual taker. In contrast, college admissions exams do have a level of consequence – admission to college of choice, potential scholarships, and certain course choice and placement options.

It was initially questioned if differences in the amounts of effort given and the levels of importance attached to test taking would be moderated by respondent's college plans (college-bound or not). College-bound respondents reported neither giving greater effort nor attaching any greater importance than did their non-college bound peers on the low-stakes state exam. In contrast, significant differences in Effort and Importance favoring college-bound occurred for college admissions testing. Although these results again showed greater Effort and Importance associated with college testing, this finding is not conclusive given the disproportion in sample sizes for these groupings (college-bound, $n = 139$; non-college bound, $n = 17$).

Grade level effects varied but not to the point of finding significant differences between grade levels within a type of testing. Interestingly, mean scores for Effort associated with college admissions increased from the 9th to the 12th grade levels while the opposite trend occurred for Importance for the state exam. An overall pattern of lower scores for effort and importance occurred for 12th graders. However, state Importance was invariant and 12th graders reported the lowest overall scores for Effort and Importance. Notwithstanding these results, college admissions scores resulted in a mean difference of + 4.51 mean score points compared to the state exam.

There were no specific sex effects, with mean scores differing fractionally between Effort and Importance for college admissions and state testing. However, a grand mean score of 21.76 (s.d. = 3.3) for college testing exceeded a grand mean score of 17.26 (s.d. = 4.3) for state tests. Again, the data support the idea that state level testing is a “low stakes” affair for the individual student when compared to college testing, which appears to be more of a motivational event. Therefore, it appears overall that the concept of low- and high-stakes status is operative in regard to the two types of tests.

CHAPTER FIVE: FINDINGS, CONCLUSIONS AND DISCUSSION

Statement of the Problem

The purpose of this study is to determine what differences exist in test-taking motivation (TTM) among high school students for completing a college admission test and the annual WV state-mandated achievement test. In the case of the admission test, the exam is not required but has future college going implications for students, including academic scholarships. The opposite is true for the state test, which is required by WV State Policy 2340 (§126-14-4.1), and searches of best practices have not produced any evidence of personal value for students beyond intrinsic value. However, state testing does have serious consequences for schools and school personnel in the No Child Left Behind compliance malaise. At issue is whether test takers are giving the same serious and genuine attention to these two different types of academic measure.

“Are high school students differing in the levels of motivation and energy when taking these exams?” “Are these students attaining an optimal score in each case and obtaining results that reflect a true or reasonably useful level of school achievement, particularly in the case of the state test?” Finally, “What differences exist in testing motivation between college bound and non-college bound students, the sex of the students, and their grade levels?” Therefore, the central purpose of the investigation was to determine the test-taking motivation among high school students regarding the levels of effort given and importance attached to testing.

Research Questions

1. What are the levels of test-taking motivation (TTM) reported by high school students who complete the WV mandatory assessment and ACT or SAT college admission tests?
2. What differences exist in TTM reported by high school students who complete the WV mandatory assessment and the ACT or SAT college admission tests?

3. What differences exist in TTM between college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?
4. What are the differences in TTM reported by college-bound high school students who complete the low-stakes WV mandatory assessment and by the same students who take the high-stakes ACT or SAT college admission tests?
5. What grade level differences exist in TTM reported by high school students who complete the low-stakes WV mandatory assessment?
6. What grade level differences exist in TTM reported by college-bound high school students who take the high-stakes ACT or SAT college admission tests?
7. What differences exist in TTM reported by high school girls and boys who complete the low-stakes WV mandatory assessment?
8. What differences exist in TTM between those reported by high school girls and boys who complete the high-stakes ACT or SAT college admission tests?

Sample and Data Collection

The sample for the study was composed of 161 high school students randomly selected from a target population of 2,969 students from five schools in a single Appalachian district. Gender (sex) responders included 75 females, 79 males, and 7 other participants who left the gender item blank. The sample consisted of 52 ninth graders, 45 tenth graders, 7 eleventh graders, and 54 twelfth graders, along with three missing values. Frequencies for 11th and 12th grades were combined for analysis. The district includes five high schools, with the following sample participation: 51 from Central High, four from County High, 20 from Mountaintop High, 41 from River City High, 44 from Valley High, and one missing value. On the survey, 143

participants indicated a college-bound plan by completing a portion of the survey that referred to a college admission test; the 18 remaining were considered non-college bound.

The data collection instrument constructed for the current study was a survey built upon the Sundre's (2002). Student Opinion Scale (SOS) student information section asked for gender (sex), grade level, and school. The survey questionnaire portion was the SOS, a test-taking motivation survey, a pre-existing instrument that consists of 10 Likert-type items assessing students' perceptions of Effort and Importance with a 5-point self-rating scale, with a minimum of 1 representing strong disagreement and a maximum of 5 representing strong agreement.

For the current study, the compound SOS survey was designed to examine the relationship between the levels of Effort expended and Importance attached to two kinds of standardized assessments that students commonly encounter. Taken together, these two concepts are assumed to yield a level of test-taking motivation.

To achieve the goals of this study, the SOS was presented in three different reference contexts: Survey #1, labelled "Main Survey," which asked participants to reflect upon the last state level exam they completed; "Survey #2," which asked participants to reflect upon a past college admission test; and "Survey #3," which asked participants to think upon a college admission test they plan to take in the future.

Conclusions

Research Question 1.

What are the levels of TTM reported by students who complete the state level exam and a college admission test?

The major finding was that respondents reported statistically significant greater levels (mean scores) for effort and importance on the high-stakes college admission tests compared to

lower mean scores on the low-stakes state level exam. Additionally, greater variability was reported for effort and importance for the state exam (low stakes). Overall, these participants appeared to respond with greater variability, perhaps being unsure about their perceptions of the state exam. Sundre (1997) found, among a sample of college students, that the “importance” dimension accounted for about 50% of the variance while effort accounted for about 20% of the variance. Abdlefatah (2010), studying the association of test-taking motivation with 9th graders, similarly noted that mean scores for importance were greater than those for effort. Results for the current study provide evidence that simply indicating that given concepts have equitable importance does not necessarily mean that equitable effort will be expended.

Research Question 2.

Effort – What differences exist in the levels of Effort reported by students regarding the low-stakes WV state level exam and the high-stakes college admission tests?

Respondents reported expending significantly greater effort when completing college admissions tests than when completing the WV mandatory assessment. Not only was the difference significant at the 95% confidence level ($p < .001$), but the effect size was quite large, $d = 1.08$, suggesting that high school students tried harder on college admissions tests than on state level exams.

Importance – What differences exist in the levels of Importance reported by students who complete the WV state level exam and college admission tests?

The significant difference in SOS importance levels supports the idea expressed previously that perceptions from students regarding the level of importance attached to college admissions testing is significantly stronger when compared to state level testing. This finding also carries a very high level of magnitude, supported statistically with an effect size $d = 1.22$.

Covington and Beery (1976) proposed that students strive to maintain a sense of “self-ability.” This student perception may manifest itself in the test-taking context when a student needs to see oneself as competent, and this need likely motivates one to expend effort toward that goal. This phenomenon may be at work for college admissions testing.

Research Question 3.

Effort – What differences exist in the levels of Effort among college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?

Overall, levels of effort were not distinguished by college or non-college plans on the state level exam. However, this outcome was likely affected by the disproportion of college bound to non-college bound respondents, with a ratio of respondents of approximately eight college bound students to one non-college bound student.

Importance – What differences exist in the levels of Importance among college-bound and non-college bound high school students who complete the low-stakes WV mandatory assessment?

Overall, the levels of importance were not distinguished by college or non-college plans on the state level exam. Again, this outcome was likely affected by the disproportion of college bound to non-college bound respondents, with a ratio of respondents of approximately eight college bound students to one non-college bound student.

Research Question 4.

Effort – What is the difference in levels of Effort reported by college-bound students who complete the low-stakes WV mandatory assessment and by the same students who take the high-stakes ACT or SAT college admission tests?

This analysis yielded a significant difference in favor of mean scores for Effort reported by college bound students who completed the college admissions tests compared to the same sample who completed the state level exam. Mean score differences resulted in an effect size measure of 1.07, favoring college tests.

Comparing this result with the Effort result of Research Question Two, which showed an effect size $d = 1.08$, it would seem that the two are closely related due directly to the disproportion of the college bound and non-college bound respondents. The two analyses were based on the responses of nearly the same sets of participants, because the college-bound students comprised most of the entire sample.

Importance – What is the difference in levels of Importance reported by college-bound students who complete the WV state level exam and by the same students who take a college admission test?

This finding yielded a significant difference in mean scores for Importance reported by college bound students who completed the college admissions tests compared to the same sample who completed the state level exam. Mean score differences resulted in an effect size measure of 1.21. This finding is similar to that found for Importance in Research Question Two that resulted in an effect size of 1.22.

Research Question 5.

What grade level differences exist in TTM reported by all students who complete the low-stakes WV mandatory assessment?

The trend was found for Effort and Importance scores to decrease slightly up the grade levels as the lowest scores, on average, occurred for 12th graders and the highest for 9th graders.

For this sample, it appears that students found the state level exam less important and gave lesser effort on it as grade level increased through high school.

Research Question 6.

What grade level differences exist in TTM reported by high school students who take the high-stakes ACT or SAT college admission tests?

Respondents equitably reported high levels of Effort and Importance for college admission tests, but with no significant difference across grade levels. For the sample it appears, again, that high school students found college admission tests quite important and put forth greater effort, regardless of grade level when compared to the state level exam.

Research Question 7.

What differences exist in TTM reported by girls and boys who complete the state level exam?

Although the finding for sex differences when analyzing for Effort on the state level exam yielded a significant difference favoring girls, this result also yielded a very small effect size of $d = .028$. Similarly, the result for Importance showed significance, again favoring girls, with a relatively small effect size of $d = .05$. Even though the mean scores resulted in statistical significance for girls, these differences most likely are not of practical importance.

Research Question 8.

What differences exist in the levels of TTM reported by girls and boys who take the high-stakes ACT or SAT college admission tests?

Both girls and boys reported similar mean scores for college admission tests, with no significant difference found. Sex was not a distinguishing variable for Effort and Importance regarding college admissions tests, and the same result occurred for boys and girls on state level

tests. However, a significant difference was found when comparing Effort and Importance between college admissions tests and state level tests.

Discussion and Implications

Conclusions

Overall, it is apparent that these high school students differ perceptually about the motivation they have for state level testing compared to college admissions testing. The intersection of the scores for effort and importance clearly shows a stronger perception for college admissions testing. And if students are not expending sufficient effort on state level testing what does this mean for the results that are being used by policy makers to assess related school accountability mandates? If low motivation (i.e., lesser effort and importance) leads to test scores that are not optimal measures of student achievement, then test validity becomes a critical issue – both educationally and politically. It then becomes essential to assess and monitor the degree of effort students expend.

NAEP has been the source of extensive research regarding test-taking motivation, including the effects of monetary incentives. Offering financial rewards had mixed results: about one-third of a national sample of 8th graders and about one-half of a national sample of 12th graders reported “somewhat” or “not at all” regarding the level of effort expended for financial rewards. About one-third of 8th graders and about two-thirds of the 12th graders reported their perceptions as “somewhat important” or “not very important” regarding the importance of financial rewards. These kinds of results have been generalized to state level testing programs (ETS, 1991; O’Neil, Sugrue, Abedi, Baker, & Golan, 1997).

Ninth graders scored the highest compared to 12th graders who scored the lowest for effort and importance on the state level exam. It appears for this grouping that the state level

exam is given lesser importance as students progress through the grades, a result that supports previous research on such grade-level effects. For example, Paris, et al. (1992) studied age-related differences in regard to test-taking motivation among 8th and 12th graders and found lesser motivation to do well occurred for older students. Kiplinger and Linn (1993) noted essentially the same for 12th graders. Conversely, the current study found that grade level differences were not a distinguishing factor in regard to student perceptions for college admissions testing for this sample. However, when compared to state level testing, greater effort and importance were associated with college admissions tests.

Scores for female students were somewhat greater but not significant statistically. They were no more likely than male respondents to give greater effort and to attach greater importance neither for the state level exam nor for college admissions tests. Whatever “stakes” were operational within the sample; gender was not a distinguishing factor. This finding is in contrast to several studies noted in the literature. O’Neil (1997) noted that males who were measured at lower levels of motivation for test taking tended to expend lesser effort. Demars, et al. (2013) confirms O’Neil, indicating that females put forth greater effort but largely due to selected personality factors.

Contribution to the Current Body of Knowledge

The Student Opinion Scale Test Manual (2007) summarizes a study comparing low- and high-stakes assessments at James Madison University (JMU). The low-stakes context included two years of SOS results from freshmen and sophomores in reference to a required set of general education assessments as part of that university’s Fall and Spring Assessment Days. The high-stakes portion was comprised of seven years of SOS results among graduating Social Work seniors in reference to motivation on their mandatory comprehensive exam in their major. The

JMU results included highly significant differences in Effort ($d = 1.30$) and Importance ($d = 2.30$). The results of the current study are comparable to those from the JMU study with similarly high effect sizes for the differences in Effort and Importance. The results of the current study indicate agreement with the JMU result and do provide related evidence that the phenomenon is evident at the high school level regarding large scale, low- and high-stakes tests.

Implications

Any result indicating low test-taking motivation questions the potential influence of low motivation on test score validity. How valid are low-stakes test scores if students report far less expended effort than on high-stakes tests? Can such test scores be trusted to be reported publicly and to influence consequential policy decisions as laid out in federal and state law? The question of test score validity is threaded through the literature since Albert Shanker in his New York Times column in 1990 inquired about the discrepancy between high-stakes Regents examination scores and low-stakes NAEP scores. Shanker publicly put forth the issue of test score validity when the test itself has no benefit or consequence to the test taker. His suggestion to conduct studies that manipulate various aspects of the issue spurred over two decades of research. Notable examples included Kiplinger and Linn (1993), who manipulated test stakes in comparison to student performance; Wolf and Smith (1995), who devised an early motivation questionnaire to gauge the influence of test consequences on self-reported motivation; Wise, Bhola, and Yang (2006) who used item response time on computer-based tests to yield a measure of test-taking effort; and Swerdzinski (2011) who investigated the effects of response time and effort.

Given the results of the current study, the issue of test validity has been shown to be relevant regarding the state level exam for a sample of high school students in West Virginia.

Because test-taking motivation has been shown to influence test performance (Eklöf, 2014), the question of the potential influence of low motivation on test score validity should not be ignored by school administration and policymakers.

State and local public school officials and administrators, in collaboration, can use the results of this study to understand why state level testing may not be a high priority among high school test takers. It seems clear that these tests are low-stakes affairs for students, although not necessarily a purposeful behavior on their part. Officials can also use the results of the study to identify specific test-taking orientation activities and/or programs, including recent computer-based programs, to enhance student motivation on state level exams. Student perception of the value of an activity is the reason for engaging in that activity. It is a choice. Eccles and Wigfield (2002) surmised that mandatory school assignments and standardized tests are not voluntary and have lesser value and expectancy for success. From the student's point of view, does anything good or bad happen in regard to the student's performance? In contrast, schools and teachers have definite consequences for high stakes test-taking results ranging from being publicized as "distinguished" to being completely taken over by outside agencies.

School officials may assume that their students have a good understanding of the importance of state testing results and its related compliance policies and expectations. However, it is quite unlikely that students are not aware of the specific educational and political ramifications for school administrators and teachers. In addition, it is quite likely that school personnel may not be aware of the test-taking motivational levels that students bring to the testing context, notwithstanding, the various reminders and activities that are in place to urge students to do their best. It is certainly questionable whether students are giving an optimal effort on test performance. Why might they not take these tests seriously? Their test scores are

not graded nor do these have any effect on grade promotion and retention, graduation, career development or placement, or scholarship awards (IASA, 1994; Stefanos & Parks, 2003; Swerdzewski, Harnes & Finney, 2011; Wise 2004). Moreover, there does not seem to be any substantial attention given by their schools to students who perform well on the exam. In effect, there is little or no consequence for a good, bad or in-between performance when compared to the results of college admission tests. This treatment is in contrast with performance on the latter type of test, which offers the opportunities of college choice, college admission, financial aid and related future incentives.

This study can provide the foundation for constructing alternative and newer strategies and/or programs to enhance student readiness and intrinsic motivation for state level exams. There is evidence that progress is possible in this area. Hawthorne, et al. (2015) found, among undergraduate students, that personalized motivational prompts had a positive influence on mean scores on a low-stakes assessment. Although their result did not include a positive influence on critical thinking subscores or on self-reported effort or importance, perhaps a similar strategy using personalized motivation prompts might be applied to high school populations regarding a low-stakes state level exam.

The current study can also highlight a need to re-assess the policy and practice of statewide testing. For example, the current results support the statewide use of certain practices found in the related literature. Because test stakes were shown to influence the effort expended on high school level tests in the current study, it may behoove state level administration and officials to attach stakes related to grade level promotion or graduation, or related incentives that could provide the student-perceived value necessary to yield more genuine effort on the state level exam.

Also, given the ubiquitous nature of achievement testing nationwide, along with previous research results, computer-based tests could be monitored for Response Time Effort, and the test results could be cleaned using motivation filtering *before* test scores are reported and consequences doled out to districts and schools. This suggestion implies using research results to address the question of test score validity by using results that have been shown to be more valid for the purposes of public reporting. In fact, the Northwest Education Association (NWEA) in the administration of its MAP test recently adopted this practice. When response time effort is measured below a certain threshold for at least 30% of test items for a student, then NWEA recommends that the entire test score be considered invalid for that student (NWEA, 2017). In addition, there is the idea of monetary rewards to students for just participating in testing and for improved test scores on classroom tests.

Recommendations for Further Research

Although the study provided useful results regarding the effects of test-taking motivation among the students in these schools, the following aspects may provide further refinement and continue to build a research database.

1. The study was built on a quantitative research design model. Adding a qualitative component such as a personal interview of selected students across the schools may provide more in-depth knowledge about the level of motivation and the seriousness that these students bring into the testing contexts.
2. A predictive analytical model that includes the existing variables (sex, college plans, and grade level) could provide specific analyses of the effects (variance) of separate predictors and also account for combined effects of variance.

3. The sample resulted in a disproportional number of those planning to attend college or not as well as for grade level distribution for the 11th grade factor. Ensuring a sample that would be proportionally (equitably) representative of these variables could likely provide greater insight into the potential effects on student perceptions of effort and importance.
4. The study did not include a demographic variable related to current student achievement. Perceptual differences in the data regarding effort and importance could be moderated somewhat by those who are high achievers and low achievers.
5. The study was based on a return of 161 from a population of about 3,000. Future studies could ensure alternative data collection strategies within the schools to obtain greater sample sizes as a whole as well as for the specific factors in the study. The use of Response Time Effort on computer-based tests may be a viable solution to obtaining large-scale data of this nature as well as estimating and monitoring rapid guessing as an indicator of low effort (Wise & Kong, 2005).
6. The study examined a one-shot sample in 2017. Extending this type of study to obtain longitudinal data per student, per school, and per region may provide insight into the potential influence of community such as changes in populations and economic conditions.
7. The study examined high school students in grades 9 - 12 in north central West Virginia. Expanding the parameters of the population to schools in northern and southern West Virginia, or even regional or nationwide, could provide more representative data to confirm the generalizations found in the study.
8. The SOS survey tool directed students to reflect upon 10 items, five each regarding the level of effort and importance for the state level exam and the college admissions test.

Expanding the survey to include additional items may extract greater refinement regarding student perceptions. This practice is supported in part by the results obtained for the last decade in Sweden by Hanna Eklöf (2006; 2014), who has included measures of test anxiety, expectancy, and interest in addition to effort and importance in her studies of test-taking motivation.

9. The study did not include a parent variable regarding the level of understanding and involvement that parents may have had in guiding and motivating their children for the state exam. Adding or expanding the role of the parent within survey items on the SOS could provide an additional perspective on the results.

REFERENCES

- Abdelfattah, F. (2010). The relationship between motivation and achievement in low-stakes examinations. *Social Behavior and Personality*, 38(2).
doi:10.2224/sbp.2010.38.2.159
- Achieve. (2008). Out of many, one: Toward rigorous common core standards from the ground up. Achieve, Washington, D.C. Retrieved from <http://www.achieve.org>
- Ahn, T., & Vigdor, J.L. (2011). Making teacher incentives work: Lessons from North Carolina's teacher bonus program. Retrieved from <http://www.aei.org>
- Appalachian Regional Commission. (2015). County economic status in Appalachia, FY 2015. Retrieved from <http://www.arc.gov/>
- Arvey, R. D., Strickland, W. Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695- 716.
- Association of Test Publishers. (2013). Member list. Retrieved from <http://www.testpublishers.org/>
- Atkinson, J. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 259-372.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9). Retrieved from <http://exordio.qfb.umich.mx>
- Barry, C.L., & Finney, S.J. (2009). Does it matter how data are collected? A comparison of testing conditions and the implications for validity. *Research & Practice in Assessment*, 4. Retrieved from <http://www.rpajournal.com>
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2). doi: 10.1037/0033-295X.113.2.409

- Brown, S.M. & Walberg, H.J. (1993). Motivational effects on test scores of elementary students. *Journal of Educational Research* 86(3).
- Capella University. (2008). Dissertation Guides Workbook – Chapters 1-5 2008-2009. School of Psychology & School of Human Services. Minneapolis, MN. Retrieved from <https://www.capella.edu>
- Center on Education Policy. (2008). A call to restructure restructuring: Lessons from the No Child Left Behind Act in five states. Retrieved from <http://www.cep-dc.org>
- Chingos, M.M. (2012). State spending on K-12 assessment systems. Brown Center on Education Policy at Brookings. Retrieved from <http://www.brookings.edu>
- Crocco, M.S., & Costigan, A.T. (2007). The Narrowing of Curriculum and Pedagogy in the Age of Accountability Urban Educators Speak Out *Urban Education* 42. Retrieved from <http://ue.sagepub.com/content/42/6/512.full.pdf+html>
- Connell, J.P. (1985). A new multidimensional measure of children's perception of control. *Child Development*, 56.
- Covington, M., & Beery, R. (1976). *Self-worth and School Learning*. New York: Holt, Rinehart, & Winston.
- Curriculum Review*. (2005). States tangle with feds over NCLB waivers. *Curriculum Review*, 44(9).
- Deci, E.L., & Ryan, R.M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum
- DeMars, C.E., Bashkov, B.M., & Socha, A.B. (2013). The role of gender in test-taking motivation under low-stakes conditions. *Research & Practice in Assessment*, 8. Retrieved from <http://www.rpajournal.com>

Doorey, N. (December 2012/January 2013). Coming soon: A new generation of assessments.

Educational Leadership 70(4). Retrieved from <http://www.coxsackie-athens.org>

Eccles (Parsons), J., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence, (Ed.), *Achievement and Achievement Motivation*. Freeman. San Francisco.

Eccles (Parsons), J., Adler, T.F., & Meece, J.L. (1984). Sex differences in achievement: A Test of alternate theories. *Journal of Personality and Social Psychology*, 46, 1. Retrieved from <http://rcgd.isr.umich.edu>

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. doi:10.1177/0146167295213003

Eccles, J. S., & Wigfield, A. (2000). Schooling's influences on motivation and achievement. In S. Danziger and J. Waldfogel, (Eds.). *Securing the future: Investing in children from birth to college*. Retrieved from <http://rcgd.isr.umich.edu/garp/articles/eccles00b.pdf>

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1). doi: 10.1146/annurev.psych.53.100901.135153

Eccles, J.S., Wigfield, A., Harold, R., & Blumenfeld, P.B. (1993). Age and gender differences in children's self-and task perceptions during elementary school. *Child Development*, 64.

Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*, 66. doi:10.1177/0013164405278574

- Eklöf, H. & Knekta, E. (2014). Different stakes, different motivation? Swedish studies of test-taking motivation in different assessment contexts. Commentary paper AERA conference.
- Fan, W., Williams, C.M, & Wolters, C.A. (2012). Parental involvement in predicting school motivation: Similar and differential effects across ethnic groups. *Journal of Educational Research*, 105(1), 21-35. doi:10.1080/00220671.2010.515625
- Feather, N.T. (1988). Values, valences, and course enrollment: Testing the role of personal values within an expectancy-value framework. *Journal of Educational Psychology*, 80.
- Fryer, R.G., Jr. (2010). Financial incentives and student achievement: Evidence from randomized trials. NBER Working Paper 15898. Retrieved from <http://www.eblofdallas.org>
- Fryer, R.G., Jr. (2011). Teacher incentives and student achievement: Evidence from New York City Public Schools. Harvard University and NBER. Retrieved from <http://scholar.harvard.edu>
- Gallagher, C.J. (2003). Reconciling a tradition of testing with a new learning paradigm. *Educational Psychology Review*, 15(1). doi:1040-726X/03/0300-0083/0
- Goldin, C. (1999). A brief history of education in the United States. *National Bureau of Economic Research Working Paper Series on Historical Factors in Long Run Growth*, Historical Working Paper No. 119. Retrieved from <http://www.nber.org>
- Gollwitzer, P. M., & Oettingen, G. (2002). Psychology of motivation and actions. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 10105-10109). Oxford, U.K.: Elsevier. Retrieved from <http://psych.nyu.edu/>

- Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner & R. C. Calfee (Eds.). Handbook of educational psychology (pp. 63-84). New York: Simon & Schuster Macmillan. Retrieved from <https://www.msu.edu>
- Gulek, C. (2003). Preparing for high-stakes testing. *Theory into Practice*, 42(1).
doi:10.1353/tip.2003.0006
- Hamilton, M., Vasquez Heilig, J. & Pazez, B. (2013). [A nostrum of school reform?: Turning around reconstituted urban Texas high schools](#). *Urban Education*.
doi:10.1177/0042085913475636
- Hanna, J. (2005). The Elementary and Secondary Education Act: 40 years later. *Harvard Graduate School of Education News* (1 June, 2005). Retrieved from <http://www.gse.harvard.edu>
- Hanson, F. A. (1993). *Testing Testing: The Social Consequences of the Examined Life*. University of California Press. Retrieved from http://books.google.com/books?id=vk7MhEF_TYAC
- Hanushek, E.A. & Raymond, M.E. (2004). Does school accountability lead to improved student performance? *National Bureau of Economic Research Working Paper Series*, Working Paper No. 10591. Retrieved from <http://www.nber.org>
- Hawthorne, K.A., Bol, L., Pribesh, S., & Suh, Y. (2015). Effects of motivational prompts on motivation, effort, and performance on a low-stakes standardized test. *Research & Practice in Assessment*, 10.
- Hoff, D.J. (2003). Nonprofit ETS making quick surge into K-12 market. *Education Week*, 22(38).
- Hoff, D.J. (2005). Va. Board seeks waivers from provisions of NCLB. *Education Week*, 24(20).

- Hull, C. (1943). *Principles of Behavior*. New York: Appleton-Century-Crofts.
- Improving America's Schools Act. (1994). Pub. L. 103-382, 1994, 108 Stat. 3518
- Karmos, A.H. & Karmos, J.S. (1984). Attitudes toward standardized achievement tests and their relation to achievement test performance. *Measurement and Evaluation in Counseling and Development*, 17.
- Kiplinger, V. L., & Linn, R. L. (1993). Raising the stakes of test administration: The impact on student performance on NAEP. (CSE Tech. Rep. No. 360). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. Retrieved from <http://www.cse.ucla.edu>
- Kress, S., Zechmann, S., & Schmitten, J.M. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation*, 48(1). Retrieved from <http://ezproxy.marshall.edu:2082/hottopics/lnacademic/?verb=sr&csi=143841>
- Kuznia, R. (2012). Some schools offer incentives to students to improve state test scores. Retrieved from <http://www.dailybreeze.com>
- Lang, J. W. B., & Fries, S. (2006). A revised 10-item version of the Achievement Motives Scale: Psychometric properties in German-speaking samples. *European Journal of Psychological Assessment*, 22, 216-224. doi: 10.1027/1015-5759.22.3.216
- Lee, J. (2010). Dual standards of school performance and funding? Empirical searches of school funding adequacy in Kentucky and Maine. *Education Economics* 18(2). doi:10.1080/09645290902796415

- Luecht, R.M., & Sireci, S.G. (2012). A Review of models of computer-based testing. The College Board. Retrieved from <https://research.collegeboard.org>
- Levitt, S.D., List, J.A., Neckermann, S., & Sadoff, S. (2012). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. NBER Working Paper No. 18165. Retrieved from <http://www.nber.org>
- Lewin, K. (1935). *A Dynamic Theory of Personality*. New York: McGraw-Hill.
- Lewin, K., Dembo, T., Festinger, L., & Sears, P. (1944). Level of aspiration. In J. M. Hunt, (Ed.), *Personality and the Behavioral Disorders*. (Vol. 1, pp. 333-378). New York: Ronald Press.
- Madaus, G. & Russell, M. (2010). Paradoxes of high-stakes testing. *Journal of Education*, 190(1/2).
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The Achievement Motive*. East Norwalk, CT US: Appleton-Century-Crofts.
- McCloskey, W. & McMunn, N. (2000). Strategies for dealing with high-stakes state tests. *Phi Delta Kappan*, 80(2).
- Mead, S. (2007). Easy way out: “Restructured” usually means little has changed. *Education Next*, 7(1). Retrieved from <http://educationnext.org>
- Meece, J.L. (1991). The classroom context and students’ motivational goals. In M.L. Maehr and P.R. Pintrich (Eds.), *Advances in Motivation and Achievement* (Vol. 7, pp. 261-285). Greenwich, CT. JAI Press.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- National Defense Education Act of 1958. Pub. L. 85-864.

- National Conference of State Legislatures. (2005). Task Force on No Child Left Behind Final Report. Retrieved from <http://www.hartfordinfo.org>.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common Core State Standards, National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C. Retrieved from <http://www.corestandards.org>
- Nicholls, J.G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3). Retrieved from <http://gribouts.free.fr>
- No Child Left Behind Act of 2002. Pub. L. 107-110. <http://uscode.house.gov>
- Northwest Education Association. (2017). Keep students engaged. Retrieved from <https://teach.mapnwea.org>.
- Olson, L. (2005). NCLB waiver lets Virginia offer tutoring before choice. *Education Week*, 25(2).
- O'Neil, H. F., Jr., Abedi, J., Lee, C., Miyoshi, J., & Mastergeorge, A. (2004). Monetary incentives for low-stakes tests (CSE Rep. No. 625). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. Retrieved from <http://www.cse.ucla.edu>
- O'Neil, H. F., Jr., Sugrue, B., Abedi, J., Baker, E. L., & Golan, S. (1997). Final report of experimental studies on motivation and NAEP test performance (CSE Tech. Rep. No.427). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. Retrieved from <http://www.cse.ucla.edu>

- O'Neil, H.F., Jr., Sugrue, B., & Baker, E.L. (1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, 3, 135-157.
- Paris, S. G., Lawton, T. A., & Turner, J. C. (1992). Reforming achievement testing to promote students' learning. In C. Collins & J. N. Mangieri (Eds.), *Teaching thinking: An agenda for the twenty-first century*. Preview retrieved from <http://books.google.com/>
Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ray, J.J. (1975). A behavior inventory to measure achievement motivation. *The Journal of Social Psychology*, 95, 135-136. Retrieved from <http://jonjayray.tripod.com/>
- Ray, J.J. (1979). A quick measure of achievement motivation -- Validated in Australia and reliable in Britain and South Africa. *Australian Psychologist*, 14, 337-344. Retrieved from <http://jonjayray.tripod.com/>
- Resnick, D.P. (1980). Chapter 1: Minimum competency testing historically considered. *Review of Research in Education*, 8(3). doi: 10.3102/0091732X008001003
- Rothman, R. (1995). *Tests of Significance*. Jossey-Bass, San Francisco.
- Creative Research Systems. (2012). Sample size calculator. Accessed at <http://www.surveysystem.com/>
- Schiefele, U. 1999. Interest and learning from text. *Scientific Studies of Reading*, 3.
- Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77(5), 629-637.
- Scholl, R.W. (2002). Motivation: Expectancy theory. Retrieved from <http://www.uri.edu/>

- Schuler, H., Thornton, G.C., Frintrup, A., Mueller-Hanson, R. (2004). *AMI: Achievement Motivation Inventory Manual*. Boston, MA. Hogrefe Publishing. Product description retrieved from <http://www.hogrefe.com/>
- Schultz, D.P. and Schultz, S.E. (2009). *Theories of Personality* (9th ed.). Belmont, CA: Wadsworth, Cengage Learning. Retrieved from <http://cengagesites.com/>
- Seligman, M. (1975). *Helplessness: On Depression, Development, and Death*. San Francisco: Freeman.
- Shanker, A. (1990). How much do our kids really know? Raising the stakes on NAEP. *The New York Times*, July 29. Retrieved from <http://source.nysut.org/>
- Stefanou, C. & Parkes, J. (2003). Effects of classroom assessment on student motivation in fifth-grade science. *The Journal of Education Research*, 96(3). Retrieved from <http://www.stanford.edu/>.
- Sundre, D.L. (1996). The role of examinee motivation in assessment: Celebrity, scene stealer, or cameo? Paper presented at 1996 AERA annual meeting. Retrieved from <http://www.jmu.edu>
- Sundre, D.L. (1997). Differential examinee motivation and validity: A Dangerous combination. Retrieved from <http://www.jmu.edu>
- Sundre, D. L. (1999). Does examinee motivation moderate the relationship between test consequences and test performance? (Report No. TM029964). Harrisonburg, Virginia: James Madison University. (ERIC Documentation Reproduction Service No. ED432588).
- Sundre, D.L. (2007). *The Student Opinion Survey (SOS): A measure of examinee motivation, test manual*. The Center for Assessment and Research Studies. Harrisonburg,

- VA. Retrieved from
http://www.jmu.edu/assessment/resources/resource_files/sos_manual.pdf
- Sundre, D.L. & Finney, S.J. (2002). Enhancing the Validity and Value of Learning Assessment: Furthering the Development of a Motivation Scale. Paper Presented at the Annual Meeting of the American Educational Research Association.
- Sundre, D.L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, (29)1.
- Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation . *Assessment Update*, 14 (1), 8–9. Retrieved from <http://www.jmu.edu>
- Sundre, D.L., & Wise, S.L. (2003). ‘Motivation Filtering’: An Exploration of the Impact of Low Examinee Motivation on the Psychometric Quality of Tests. Paper presented at the National Council on Measurement in Education Annual Conference. Chicago, Illinois. Retrieved from <http://www.jmu.edu/>
- Swerdzewski, P.J., Harmes, C.J., & Finney, S.J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education*, 24. doi: 10.1080/08957347.2011.555217
- Thelk, A.D., Sundre, D.L., Horst, S.J., & Finney, S.J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *JGE: The Journal of General Education*, 58(3). The Pennsylvania State University, University Park, PA. doi: 10.1353/jge.0.0047
- U.S. Department of Education. (1983). A Nation at Risk: Findings. Retrieved from <http://www.mat.uc.pt>

- U.S. Department of Education. (2010a). Race to the Top Assessment Program. Retrieved from <http://www2.ed.gov>
- U.S. Department of Education. (2010b). Race to the Top overview. *75 Federal Register* (14 April, 2010). Retrieved from <http://www2.ed.gov>
- U.S. Department of Education. (2012). Overview: The federal role in education. Retrieved from <https://www2.ed.gov/about/overview/fed/role.html>
- U.S. Department of Education, National Center for Education Statistics. (2004). Education longitudinal study, 2002: Base year. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR04275.v1
- Veroff, J., McClelland, L., & Marquis, K. (1971). Measuring intelligence and achievement motivation in surveys. Survey Research Center: Institute for Social Research. University of Michigan. Retrieved from <https://psidonline.isr.umich.edu>
- Vinovskis, M.A. (1998). Overseeing the nation's report card: The creation and evolution of the National Assessment Governing Board (NAGB). Retrieved from <http://www.nagb.org/content/nagb/assets/documents/publications/95222.pdf>.
- von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press. Princeton, NJ.
- Vroom, V.H. (1964). *Work and Motivation*. Wiley. New York.
- Webster, D. S. (1984). The Bureau of Education's suppressed rating of colleges, 1911–1912. *History of Education Quarterly*, 24(3).
- Weiner, B. (2013). Little-known truths, quirky anecdotes, seething scandals, and even some science in the history of (primarily achievement) motivation. *Personality and Social Psychology Review*, (17)3.

- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R.M. (1971). *Perceiving the Causes of Success and Failure*. New York: General Learning Press.
- West Virginia Department of Education. (2012/2013). State report card data. Retrieved from <http://wveis.k12.wv.us/>
- West Virginia Education Association. (2009). WVBE takes over Grant County Schools. Retrieved from <http://www.wvea.org/>
- West Virginia Legislature. (2014a). School takeover fairness bill. *Wrap-Up: The Official Newsletter of the West Virginia Legislature*, 25(5). Retrieved from <http://www.legis.state.wv.us/>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1). doi:10.1006/ceps.1999.1015
- Wise, S.L., Bholá, D.S., & Yang, S. (2006). Taking the time to improve the validity of low-stakes tests: The effort-monitoring CBT. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco. Retrieved from <http://www.jmu.edu/>
- Wise, S. L., & DeMars, C. E. (2003). Examinee Motivation in Low-Stakes Assessment: Problems and Potential Solutions. Paper presented at the annual meeting of the American Association of Higher Education Assessment Conference, Seattle, WA. Retrieved from <http://www.jmu.edu/>
- Wise, S. L., & DeMars, C. E. (2010). Examinee Noneffort and the Validity of Program Assessment Results. *Educational Assessment*, 15(1). doi:10.1080/10627191003673216
- Wise, S.L., Kingsbury, G.G., Thomason, J., & Kong, X. (2004). An Investigation of Motivation Filtering in a Statewide Achievement Testing Program. Paper Presented at the annual

- meeting of the National Council of Measurement in Education, San Diego, CA. Retrieved from <http://www.jmu.edu>
- Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8 (3), 227-242.
- Woodrum, A. (2004). State-mandated testing and cultural resistance in Appalachian schools: Competing values and expectations. *Journal of Research in Rural Education*, 19(1). Retrieved from <http://jrre.psu.edu>
- Yoon, K. S., Eccles, J. S., & Wigfield, A. (1996). Self-concept of ability, value, and academic achievement: A test of causal relations. In *annual meeting of the American Educational Research Association, New York*. Retrieved from <http://www.rcgd.isr.umich.edu/garp/presentations/eccles96o.pdf>
- Zehr, M. A. (2005). California NCLB waiver denied. *Education Week*, 24(37).

APPENDICES

APPENDIX A: IRB APPROVAL LETTER



Office of Research Integrity
Institutional Review Board
One John Marshall Drive
Huntington, WV 25755

FWA 00002704

IRB1 #00002205

IRB2 #00003206

September 19, 2016

Samuel Securo, Jr., Ed.D.
Elementary and Secondary Education

RE: IRBNet ID# 892777-1

At: Marshall University Institutional Review Board #2 (Social/Behavioral)

Dear Dr. Securo, Jr.:

Protocol Title: [892777-1] A comparison of self-perceived test-taking motivation on large scale high- and low-stakes tests

Expiration Date: September 19, 2017

Site Location: MUGC

Submission Type: New Project APPROVED

Review Type: Expedited Review

In accordance with 45CFR46.110(a)(7), the above study and informed consent were granted Expedited approval today by the Marshall University Institutional Review Board #2 (Social/Behavioral) Chair for the period of 12 months. The approval will expire September 19, 2017. A continuing review request for this study must be submitted no later than 30 days prior to the expiration date.

This study is for student Randall King.

If you have any questions, please contact the Marshall University Institutional Review Board #2 (Social/Behavioral) Coordinator Bruce Day, ThD, CIP at 304-696-4303 or day50@marshall.edu. Please include your study title and reference number in all correspondence with this office.

APPENDIX B: DISTRICT-LEVEL PERMISSION

RE: Request for approval for dissertation research

Mark Manchin <mmanchin@k12.wv.us>

Tue 4/14/2015 3:28 PM

To: Randall Joseph King <king226@live.marshall.edu>;

Categories: Green Category

Looks great!! Looking forward to reviewing your completed research. Looks like it can be used by school administrators.

Good luck, Randall

From: Randall Joseph King [mailto:king226@live.marshall.edu]
Sent: Tuesday, April 14, 2015 2:42 PM
To: Mark Manchin
Cc: Patty Britton
Subject: Request for approval for dissertation research

Dr. Manchin,

you and I spoke in February regarding my survey of student test-taking motivation, and I certainly appreciated your pledge of support at that time.

Attached are the details to reacquaint you with the details of my study (which have changed since we spoke).

If you still find my study acceptable, please express your approval in a response email that I may present as evidence to my Doctoral Committee and Institutional Review Board.

If you have any questions or comments, I will be glad to address them.

Thank you,
Randall J. King

APPENDIX C: SCHOOL LEVEL PERMISSIONS

All Harrison County middle and high schools
 Are invited to participate in a Survey of Test-Taking Motivation
 Fall 2015

Researcher: Randall King, doctoral candidate
 Marshall University Graduate College
 Advisor: Dr. Samuel Securro

Please sign in agreement to include your school in this survey.

School	Administrator
Bridgeport Middle	<i>Carole Crawford</i>
Lincoln Middle	<i>Lori Scott</i>
Mountaineer Middle	<i>John Ray</i>
South Harrison Middle	<i>John Ray</i>
Washington Irving Middle	<i>Steve B. Smith</i>
Bridgeport High	<i>Mark DeFazio</i>
Liberty High	<i>James A. Kellogg</i>
Lincoln High	<i>James Lopez</i>
Robert C. Byrd High	<i>D. S. Smith</i>
South Harrison High	<i>John Moore</i>

Thank you,

Randall J. King

APPENDIX D: INITIAL TEACHER INVITATION

Dear Teacher: (name)

Hello, my name is Randall King; I teach mathematics at Liberty High and I'm working on my dissertation research project at Marshall University Graduate College. I've contacted you because you are in the right position to help with this project.

The project is a comparison of student test-taking motivation on standardized tests, and data collection consists of the completion of a student survey. Dr. Manchin has personally supported this project, and the principal at your school has agreed to participate.

Your homeroom has been randomly selected to be included in this study, and I would certainly appreciate if you would agree to participate. Before you decide, here are the basic requirements of ethical research regarding minors and a brief description of the process.

- Parental consent must be obtained (except for those 18 years of age or older)
- Participation must be voluntary, with no direct or implied coercion for students to participate
- ***Only students who return an affirmative parental consent may participate or who are 18 years of age or older and agree to do so.***

The process for each of the participating high schools is as follows. Upon your agreement, I (Randall King, Co-Investigator) will arrange a schedule with you and the school administrator for each of the selected classes at which time and date the surveys will be completed. You will then be sent paper copies of the parental consent form to distribute in your homeroom, with an agreed upon return date and deadline to keep the process short. By the end of the deadline, I will personally visit the classroom at these prescribed times and dates to collect the assents and to administer the survey to those who have been given consent. At that time all surveys will be collected. Because some teachers have regular access to computer labs, there will be a choice whether to use paper copies or the online version.

Our goal is to conduct the process with minimal interruption and inconvenience for school personnel and students. If you have any questions or concerns please contact me, Randall

King, at 304-624-3325, extension 6862 (weekdays from 12-1pm) or my supervisor, Dr. Samuel Securo, Jr. Marshall University, 304-746-8948.

Please respond to this email with “agree paper,” “agree online,” “agree both,” or “disagree” in the subject line.

Thank you for your time

Randall J. King (email: king226@live.marshall.edu)

APPENDIX E: TEACHER INSTRUCTIONS

Dear Teacher: (name)

Thank you for agreeing to participate in my study on test-taking motivation. The project (survey) will be implemented in 5 high schools, including your particular school. At each of the participating high schools, the co-investigator (Randy King) will arrange with you and the school administrator a schedule for each of the selected classes at which time and date the surveys will be completed. The co-investigator will personally visit the classroom at these prescribed times and dates to collect these materials.

Initially, parental consent forms will need to be given to students to take home for those who are less than 18 years of age. Please emphasize to students that the participation is completely voluntary, and then collect the signed forms, place in the provided packet, and maintain this in a secure location until collected by the co-investigator.

The co-investigator will personally arrive at the school site at the scheduled time and date for each school (these will vary) to administer the surveys and collect all materials. He will distribute a copy of the survey or survey link to those who have returned a parental consent agreeing for their child to participate and also to those who are 18 or older who have agreed to participate (complete the survey). He will reiterate to the students that participation is voluntary and will let them know if there are those who initially agreed to or had parental permission to participate but have now decided not to, they can choose not to complete the survey with no consequence.

Please contact me if you have questions about any part of the process, and feel free to discuss with your school principal. You may contact Randall King at 304-624 624-3325, extension 6862 or Dr. Samuel Securo, Jr. Marshall University, 304-746-8948.

FYI: your participation will be anonymous as well. The dissertation results will not include any location or participant identifiers. Thank you again for your participation, and I welcome any feedback regarding the logistics from your end.

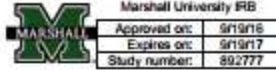
Randall J. King (king226@live.marshall.edu)

APPENDIX F: PARENTAL CONSENT LETTER

Page 1 of 2

Parental Consent/Permission **Survey of test-taking motivation**

Samuel Securo, Ed. D Principal Investigator



Introduction

Your child is invited (with your permission) to be in a research study. Research studies are designed to gain educational knowledge that may help other people in the future. There are no risks, benefits, or costs associated with being part of this research study. Participation is voluntary so please take your time to make your decision, and feel free to contact the investigator who will address any questions or concerns you may have about the project.

Why Is This Study Being Done?

The major purpose of this study is to determine how important standardized tests are to students and how much effort students put forth when completing these tests.

How Many Will Take Part In The Study?

About 3500 middle and high school students in Harrison County will take part in this study.

What Is Involved In This Research Study?

Participation in this study means filling out a brief survey about student motivation to succeed on standardized tests.

How Long Will The Survey Take?

The survey may take about 10 minutes.

Once your child begins the survey, you or your child can decide to stop participation at any time. If you or your child decide to stop participation, simply do not submit a completed survey.

What About Confidentiality?

This survey is anonymous. We will do our best to make sure that your child's personal information is kept confidential. However, we cannot guarantee absolute confidentiality. Federal law says we must keep your child's study records private. Nevertheless, under unforeseen and rare circumstances, we may be required by law to allow certain agencies to view your child's records. Those agencies would include the Marshall University IRB, Office of Research Integrity (ORI) and the federal Office of Human Research Protection (OHRP). This is to make sure that we are protecting your child's rights and safety. If we publish the information we learn from this study, your child will not be identified by name or in any other way.

Initial _____

What Are Your Rights As A Research Study Participant?

Taking part in this study is voluntary. You may choose not to allow your child to take part or you may withdraw them from the study at any time. Refusing to participate or leaving the study will not result in any penalty or loss of benefits to which you or your child are entitled.

Whom Do You Call If You Have Questions Or Problems?

For questions about the study, contact the study investigator, Randall King, at 304-624-3325 ext.6862 from 3pm – 4pm, or Dr. Samuel Securo, Jr. 304-746-8948 (securo@marshall.edu)

For questions about your rights as a research participant, contact the Marshall University IRB#2 Chairman Dr. Stephen Cooper at (304) 696-7320. You may also call this number if:

- o You have complaints about the research.
- o The research staff cannot be reached.
- o You want to talk to someone other than the research staff.

You will be given a signed and dated copy of this consent form.

SIGNATURES

You grant permission for your child _____ to take part in this study. You have had a chance to ask questions about this study and have had those questions answered. By signing this consent form you are stating that you are not giving up any legal rights to which you or your child are entitled.

Parent Name (Printed)

Parent Signature

Date

Randall King

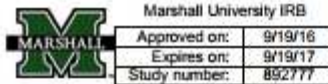
Person Obtaining Consent (Printed)

Person Obtaining Consent Signature

Date

Initial _____

APPENDIX G: STUDENT ASSENT FORM



Dear Student,

You are invited to participate in a research project entitled "Survey of Test-Taking Motivation" because you have taken standardized tests, and we would like to know some things about how you take those tests.

The study is being supervised by Dr. Samuel Securro, Jr., and Randall King will be the investigator, both from Marshall University Graduate College. This research is being conducted as part of the doctoral dissertation requirements for Randall King.

This survey consists of questions about your attitudes towards taking standardized tests and may take about 10 minutes to complete.

Your replies will be anonymous, so do not put your name anywhere on the form.

If you are under 18 years of age, this study was explained to your parents, and they gave permission for you to be in it if you want. You can talk this over with them before you decide.

If you are 18 years of age or older, you may decide for yourself to participate in this study.

There are no known benefits or risks involved with this study.

Participation is completely voluntary and there will be no penalty or loss of benefits if you choose to not participate in this research study or if you stop in the middle of the survey.

If you choose not to participate you may either return the blank survey or you may discard it.

You may choose to not answer any question by simply leaving it blank.

If you have the paper survey, then returning the completed survey to the research investigator (Randy King) means that you agree the researchers may use the answers you give.

If you are taking the survey online, then submitting the completed survey means that you agree the researchers may use the answers you give.

If you have any questions about the study, you may contact Randall King at 304-624 624-3325, extension 6862 or Dr. Samuel Securro, Jr. Marshall University, 304-746-8948.

If you have any questions concerning your rights as a research participant, you may contact the Marshall University Office of Research Integrity at (304) 696-4303.

Please keep this page for your records.

APPENDIX H: STUDENT OPINION SURVEY

STUDENT INFORMATION (Circle your information)

Gender: F M

Grade: 9 10 11 12

School: BHS LIB LIN RCB SHHS

SURVEY

- **Please think about the last standardized test you took in school (for example, the Smarter Balanced Assessment).**

Use the following rating scale:

1=Strongly Disagree 2=Disagree 3=Neutral 4=Agree 5=Strongly Agree

For each statement, circle your choice that best describes how much you agree.

- | | | | | | |
|--|---|---|---|---|---|
| 1. Doing well on this test was important to me. | 1 | 2 | 3 | 4 | 5 |
| 2. I engaged in good effort throughout this test. | 1 | 2 | 3 | 4 | 5 |
| 3. I am not curious about how I did on this test relative to others. | 1 | 2 | 3 | 4 | 5 |
| 4. I am not concerned about the score I receive on this test. | 1 | 2 | 3 | 4 | 5 |
| 5. This was an important test to me. | 1 | 2 | 3 | 4 | 5 |
| 6. I gave my best effort on this test. | 1 | 2 | 3 | 4 | 5 |
| 7. While taking this test, I could have worked harder on it. | 1 | 2 | 3 | 4 | 5 |
| 8. I would like to know how well I did on this test. | 1 | 2 | 3 | 4 | 5 |
| 9. I did not give this test my full attention while completing it | 1 | 2 | 3 | 4 | 5 |
| 10. While taking this test, I was able to complete it. | 1 | 2 | 3 | 4 | 5 |

✧If you plan to attend college, please proceed to **SURVEY #2**.

❖ **If you DO NOT plan to attend college:** thank you for your time! If you want your answers to be part of this research project, please return your completed survey to your teacher.

SURVEY #2

❖ If you have **ALREADY TAKEN** a college admissions test (ACT or SAT), please complete the survey below.

❖ If you have **NOT YET TAKEN** a college admissions test (ACT or SAT) **BUT PLAN TO DO SO**, skip ahead to **SURVEY #3**.

➤ **Please think about the last ACT or SAT you took.**

Use the following rating scale:

1=Strongly Disagree 2=Disagree 3=Neutral 4=Agree 5=Strongly Agree

For each statement, circle your choice that best describes how much you agree.

- | | | | | | |
|--|---|---|---|---|---|
| 1. Doing well on this test was important to me. | 1 | 2 | 3 | 4 | 5 |
| 2. I engaged in good effort throughout this test. | 1 | 2 | 3 | 4 | 5 |
| 3. I am not curious about how I did on this test relative to others. | 1 | 2 | 3 | 4 | 5 |
| 4. I am not concerned about the score I receive on this test. | 1 | 2 | 3 | 4 | 5 |
| 5. This was an important test to me. | 1 | 2 | 3 | 4 | 5 |
| 6. I gave my best effort on this test. | 1 | 2 | 3 | 4 | 5 |
| 7. While taking this test, I could have worked harder on it. | 1 | 2 | 3 | 4 | 5 |
| 8. I would like to know how well I did on this test. | 1 | 2 | 3 | 4 | 5 |
| 9. I did not give this test my full attention while completing it | 1 | 2 | 3 | 4 | 5 |
| 10. While taking this test, I was able to complete it. | 1 | 2 | 3 | 4 | 5 |
-

Thank you for your time! If you want your answers to be part of this research project, please return your completed survey to your teacher.

SURVEY #3

➤ **Please think about an ACT or SAT you may take in the future.**

Use the following rating scale:

1=Strongly Disagree 2=Disagree 3=Neutral 4=Agree 5=Strongly Agree

For each statement, circle your choice that best describes how much you agree.

- | | | | | | |
|--|---|---|---|---|---|
| 1. Doing well on this test will be important to me. | 1 | 2 | 3 | 4 | 5 |
| 2. I will engage in good effort throughout this test. | 1 | 2 | 3 | 4 | 5 |
| 3. I will not be curious about how I do on this test relative to others. | 1 | 2 | 3 | 4 | 5 |
| 4. I will not be concerned about the score I receive on this test. | 1 | 2 | 3 | 4 | 5 |
| 5. This will be an important test to me. | 1 | 2 | 3 | 4 | 5 |
| 6. I will give my best effort on this test. | 1 | 2 | 3 | 4 | 5 |
| 7. While taking this test, I may not work as hard as I can on it. | 1 | 2 | 3 | 4 | 5 |
| 8. I would like to know how well I do on this test. | 1 | 2 | 3 | 4 | 5 |
| 9. I will not give this test my full attention while completing it | 1 | 2 | 3 | 4 | 5 |
| 10. When I take this test, I should be able to complete it. | 1 | 2 | 3 | 4 | 5 |
-

Thank you for your time! If you want your answers to be part of this research project, please return your completed survey to your teacher.

APPENDIX I: CURRICULUM VITA

316 Ceasar Place
Hilton Head, SC 29926
304-476-5123
king226@live.marshall.edu
randyking0@gmail.com

Education

Doctoral studies in Education, 2007-present, 60+ credit hours completed
Marshall University Graduate College
Charleston, WV

Master of Arts in Education, December 2003
Salem-Teikyo University
Salem, WV

Bachelor of Science in Mathematics, May 1997
Salem-Teikyo University
Salem, WV

Master's Thesis

The effect of teacher absence on student achievement
Salem-Teikyo University
2003

Dissertation (in progress)

A comparison of self-perceived test-taking motivation on large scale high- and low-stakes tests

Honors

Summa cum laude 1997, S-TU Presidential Award 1997

Professional Experience

Mathematics Teacher (currently)
Beaufort County Schools
Bluffton, SC

Harrison County Schools
Clarksburg, WV
Mathematics Teacher (2001-2017)
Substitute Teacher (1997-2001)

Adjunct Faculty: College Algebra (Fall 1998)
Salem-Teikyo University
Salem, WV

Substitute Teacher (1997-1998)
WV Industrial Home for Youth
(Former maximum security facility for youth offenders)
Salem, WV

Research Interests

I have specific interests in curriculum design regarding theory, content, organization, and pacing. I also have broad interests in teacher education.

Professional Development

West Virginia Department of Education Transition Math for Seniors Training, 2011

Carnegie Learning Systems Training, 2011

West Virginia Department of Education Algebra Support Training, 2010

Thinkfinity TRN-2 training, 2008

Curriculum Mapping workshop, Harrison County secondary math, 2007

Co-Inquiry in Mathematics workshop and subsequent work sessions, 2007

Math Field Day test development team, Harrison County, 2006

Advanced Placement training: Calculus AB, 2004

Presentations and Trainings

Thinkfinity End-User Partial Training (TCH-P), 2008 (rating: 4.98 out of 5)

Guest speaker at WV Governor's Honors Academy: Scientific and mathematical discussion of domes, 2003-2007

References

Dr. Donna Hage, NBCT, Assistant Superintendent, Harrison County Schools 304-624-3300

Kathy Marino, Math Department Chair, Liberty High, Harrison County Schools 304-624-3264