TEST ENVIRONMENT FOR OPTIMAL PERFORMANCE IN HIGH SCHOOL STUDENTS: MEASURE DEVELOPMENT AND THE RELATIONSHIP WITH STANDARDIZED TEST SCORES

A dissertation submitted to the

Kent State University

College of Education, Health, and Human Services

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

By

Randal J. Rair

August, 2019

A dissertation written by

Randal J. Rair

B.S., Youngstown State University, 1999

B.S., Youngstown State University, 2002

M.Ed., Ashland University, 2008

Ph.D., Kent State University, 2019

Approved by

__________________________, Director, Doctoral Dissertation Committee
Aryn C. Karpinski

__________________________, Member, Doctoral Dissertation Committee
Jason Schenker

__________________________, Member, Doctoral Dissertation Committee
Anthony Vander Horst

Accepted by

__________________________, Director, School of Foundations, Leadership, and Administration
Kimberly Schimmel

__________________________, Dean, College of Education, Health and Human Services
James C. Hannon

ABSTRACT

Randal J. Rair, June 2019 EVALUATION AND MEASUREMENT

TEST ENVIRONMENT FOR OPTIMAL PERFORMANCE IN HIGH SCHOOL STUDENTS: MEASURE DEVELOPMENT AND THE RELATIONSHIP WITH STANDARDIZED TEST SCORES (294 pp.)

Director of Dissertation: Aryn C. Karpinski, Ph.D.

The rise of mobile technology and increased school accountability has enmeshed the last two generations of students in unprecedented educational experiences. This has fostered learners who are varied in their communication and academic capabilities. Simultaneously, the number of standardized tests, both mandated and voluntary, has been steadily increasing. Few studies are available that examine the preferred environmental conditions for students during these tests. The current study consists of two parts: (1) the Pilot Study and (2) the Main Study. The Pilot Study examined the psychometric properties of a measure called the Test Environment for Optimal Performance (TEOP), which was created to explore the testing preferences of the current generation of college students (i.e., Millennials). Following this quantitative section of the Pilot Study, selections of Millennials were qualitatively interviewed on topics related to test environment preferences to explore the underlying factor structure meaning. Findings from both the quantitative and qualitative portions suggest that students have a partiality for physical (i.e., “Action”) and auditory (i.e., “Sound) activities when testing.

The Main Study extended the Pilot Study by investigating current high school students' (i.e., Generation Z) preferred standardized testing conditions and if these preferred conditions have a predictive relationship with high-stakes standardized test scores (i.e., the ACT). The Main Study analyzed data using a Confirmatory Factor Analysis (CFA) to examine the psychometric properties of the TEOP in the new population, and to provide further evidence to support the "Action" and "Sound" test environment factor structure. Following the CFA, Hierarchical Multiple Linear Regressions were conducted to examine the hypothesized relationships between the TEOP factors "Action" and "Sound" and high-stakes aptitude outcomes (i.e., ACT scores). The results warranted an additional post hoc Hierarchical Multiple Linear Regression using Grade Point Average (GPA) as the outcome.

The results from the Main Study advanced the outcomes from the Pilot Study by providing additional validation support for the TEOP scores in a population of high school students. The study, therefore, provided validity and reliability evidence of the TEOP for two subpopulations (i.e., Millennial college students and Generation Z high school students). The results also showed that while the TEOP scores did not have a significant predictive relationship with ACT scores, there was a significant relationship between the TEOP scores and GPA. These findings may be useful to various stakeholders looking to address students' test preparation and academic performance. The TEOP scores can be used to inform students, parents, and school administrators of the potential congruence or incongruence between students' practices and preferences while studying and the actual test environments in order to maximize their performance. Results from this correlational study should be interpreted with caution; however, future research may consider how to use the TEOP as a tool for in-class assessment preparation.

ACKNOWLEDGEMENTS

This document is dedicated to everyone who has ever helped me, there are too many of you to name.

"Those five years are going to go by whether you finish or not"
– Msgr. Robert Siffrin

"Those five years are going to go by whether you finish or not" are the words Msgr. Robert Siffrin told me when I mentioned to him that I was starting my PhD and that it would take approximately five years. Well it has been six months longer than my original estimate, but I have frequently repeated this simple statement as the responsibilities of completing my classes, comprehensive exams, and this dissertation while trying to fulfill my obligations to my job and family at times seemed overwhelming. I thank you for this inspiration you gave in a moment that you most likely have long forgotten.

To my parents and sister, I want to thank you for a lifetime of love and support. I hope this makes up for me not getting in National Honor Society :).

I am eternally grateful to my lovely wife, Sara, for her constant support and sacrifice over these last five years. In particular, thank you for fixing all my formatting SNAFU's and my whining about why Word just won't work. I love you.

Thank you to all my teachers, the good and the bad. I have learned from all of you, but especially Matt Giambattista, Sr. Regina Rogers, and Sharon Machuga. Miss

Machuga teaching me that the fish < eats the larger number has been particularly helpful as I matriculated through the Evaluation and Measurement program.

To my committee members, Dr. Jason Schenker and Dr. Anthony Vander Horst and Graduate Faculty Representative Dr. Richard Ferdig, thank you. Thank you for putting in the time and effort and thank you for adding your expertise to my dissertation. I want to wish an additional thank you to Dr. Schenker for how well you taught the Statistics and Quantitative Research classes that formed the foundation of my success in this program.

Thank you to all of my Evaluation and Measurement classmates for your support, patience, and friendship. Please know that if I can do this so can all of you. I want to thank my direct supervisors, first Dr. Nicholas Wolsonovich and then Mary Fiala for their support and understanding. I also would like to thank whoever invented the synonym feature on Microsoft Word for allowing me to appear to have a much bigger vocabulary than I do.

Words cannot express how appreciative I am of what my advisor Dr. Aryn Karpinski has done for me. You were tireless as a guide and mentor, and your demand for excellence made me a better person. In the Mount Rushmore of my life, Dr. Aryn Karpniski gets a sculpture.

Lastly thank you to my dogs and two best friends, Gunner and Erie, thank you for always being happy to see me, thank you for reminding me when it was time to take a break, and I am sorry I scared you every time I yelled at the computer.

TABLE OF CONTENTS

PART II

CHAPTER

LIST OF FIGURES

LIST OF TABLES

PART I

CHAPTER I: INTRODUCTION (PILOT STUDY)

Present day college and high school students live in a markedly different world compared to their parents and grandparents. This world includes profound differences in the way today's students live, socialize, and learn. An abundance of technology, an explosion of social media, and a heightened emphasis on differentiated and individualized instruction in classroom have created a near-perfect storm. In schools, widespread diversity has replaced the homogeneity of the past; however, in many contexts, the processes involved in testing and assessment have not changed with evolving trends (Subban, 2006). Young adults live in a world of continual sensory stimulation, yet the rules for proctoring and taking exams have remained untouched for decades. Simultaneously, the emphasis placed on standardized assessment for both students and schools has never been greater.

Educational assessment is the systematic practice of using and documenting observed data, on knowledge, skill, attitudes, and beliefs to measure and improve student learning (Allen, 2004). The term assessment is often used interchangeably with test, and the process of assessment can address the individual learner, learning community (i.e., a class, a workshop, or other organized learning groups), an academic program, an institution, or the educational system as a whole (Brookhart & Nitko, 2015). One major category of assessment includes standardized tests, some of which are high stakes. High stakes tests are any assessments that have major consequences for passing or failing (i.e., college acceptance, high school graduation (Brookhart & Nitko, 2015). More explicitly, "high stakes" are the costs of the outcome, and not a characteristic of the test itself (Olson & Sabers, 2008). Thus, the goal of these tests is to accurately portray a student's optimal performance with minimal influence from other factors such as the test environment and/or conditions.

Statement of Problem

The rise of mobile technology and increased school accountability has enmeshed college students in unique educational experiences. These experiences have fostered a generation of Millennials who are distinct in their communication and academic capabilities. Meanwhile, the number of standardized tests, both mandated and voluntary, has been steadily increasing. A dearth of current research is available on the preferred environmental conditions for testing. This Pilot Study examines the psychometric properties including the factor structure of a measure created to explore the testing preferences of the current generation of college students (i.e., Millennials). Following the quantitative measure development, a selection of Millennials were qualitatively interviewed on topics related to test environment preferences to explore the underlying factor structure meaning.

The Pilot Study will encompass two phases. Phase I (Quantitative) is comprised of a population of current traditional university students. The participating students had all taken a minimum of one standardized test as a measurement for college aptitude. Phase II (Qualitative) consists of recent high school graduates. The contributors had all taken the ACT to gain acceptance into college.

The main research question in the Pilot Study Phase I (Quantitative) was, "What are the psychometric properties (i.e., content and construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a university student population?" The following primary research question guided the Pilot Study Phase II (Qualitative), "What are recent high school graduates' perceptions of their experiences with current standardized testing practices and environments?"

Standardized Testing and Tests

The number of standardized tests required of school children in the United States (U.S.) has increased in the last two decades. While standardized tests have been a "right of passage" for years, the current escalation has been linked to the 2001 passage of the No Child Left Behind Act (Layton, 2015). The average child takes 112 mandated, standardized tests between Pre-Kindergarten and High School graduation (Hart et al., 2015). This equates to the typical student taking eight standardized tests per year. This number does not include optional tests, diagnostic assessments (e.g., to diagnose a disability), or teacher-designed classroom examinations.

The greatest testing demands are in eighth grade. During the 2014/2015 school year, in the final year of junior high school, students spent roughly 4.22 days taking standardized tests (Hart et al., 2015). This is approximately equal to 2.34% of their time in school (Layton, 2015) and does not include class preparation devoted to mandated assessments. Coupled with optional tests such as Advanced Placement (AP) and college readiness exams (e.g., the ACT and SAT), standardized test-taking has become a "way of life" for the modern U.S. student and educator.

Several, major standardized tests are commonly taken by U.S. high school students that measure academic aptitude and college readiness. Of these, the ACT is the most popular (ACT, 2017). The ACT was created in 1959 by University of Iowa Registrar Ted McCarrel and Professor Everett Lindquist as a competitor to the Scholastic Aptitude Test (SAT). The ACT was originally short for American College Test, but since 1996, has been known as the ACT (Jacobsen, 2018). In 1959, the SAT was given at elite institutions and the ACT founders saw a need for a test aimed at both acceptance to and placement within public colleges and universities. Lindquist wanted the test to measure acquired knowledge not simply logical aptitude (Edwards, 2015). The defined purpose of the ACT is to measure a student's educational development and ability to do college-level work (ACT, 2017). In addition to being an assessment of college readiness, the ACT has become part of required standardized testing in many states (Edwards, 2015). Twenty-five states require high school students take either the ACT or SAT as a graduation requirement (Gewertz, Which states require student's to take the SAT or ACT?, 2017). Sixteen of the 25 states require the ACT, and three accept either standardized test.

Over two million students took the ACT in 2017, which was 60% of that graduating class (Gewertz, 2017). Every four-year university and college in the U.S. accepts the ACT (Marklein, 2007) The ACT consists of four sections English, Reading, Math, and Science. There is also an optional writing section. Each section is scored from

1 to 36, with a composite score created from all the sections (ACT, 2017). The ACT begins at 8:00 AM and is completed by 12:15 PM and 1:15 PM if the student takes the optional writing section (ACT, 2017).

Table 1

Summary of ACT National 2017

Section	Number of Items	Time in Minutes	Average Score	College Readiness Benchmark
English	75	45	20.3	18
Reading	40	35	21.4	22
Math	60	60	20.7	22
Science	40	35	21.0	23
Composite	215	175	20.8	-

Note. Data obtained from ACT (2017).

A compendium of rules are delineated when taking the ACT to safeguard against cheating and to ensure that all test-takers have nearly equivalent test-taking experiences. Among these guidelines, the testing environment must have adequate lighting and separation from distracting noises or activities. All devices that can emit audible sounds are prohibited, and food/drinks are not permitted during testing. Additionally, students cannot select their desk/chair/set, all desks/chairs/seats must face the same direction with at least four feet in between the others, with the test beginning in the morning and ending before lunch, and test takers must all complete the same section (one at a time) and cannot work ahead or return to previous sections (ACT, 2017). Students with learning and/or physical disabilities (as two examples) can take the ACT with accommodations (ACT, 2017). Similar rules are applied to other standardized tests such as the SAT and AP examinations (The College Board, 2016).

Millennials

Unprecedented technological changes have had an impact on the personalities and educational preferences of Millennial college students. Millennials are the generation born between 1981 and 1996 (Dimcock, 2018). Technology is a natural part of the environment for Millennials, and mobile devices allow them to remain in constant connectivity wherever they travel (Oblinger, 2008). This cohort of current college students has a range of choice, more than any other generation in history. Through websites like YouTube, Netflix, and Amazon, all available on portable smartphones accompanying their every move, a song, TV show, or movie is instantly available to them. From a young age, Millennials have been able to “Google” a question from nearly any location on the planet. This is in contrast to earlier generations whose entertainment options where limited to what was on the television or radio, and even older cohorts that were limited to books currently available in the home.

The Millennial generation of children and young adults has an advanced relationship with technology that was formed from birth (Beastall, 2008). Millennials familiarity with and reliance on technology is a product of their lives being continuously broadcast on the Internet (Prensky, 2001). Young adults today, due to this lifelong relationship with technology, have an innate aptitude for multitasking, which was not present in former generations (Prensky, 2003). A description of this cohort (i.e., Millennials) and how they learn is presented in the following paragraphs.

Millennials, Learning, and Context-Dependent Memory

The formal schooling of Millennials has included a rise in the Internet and wireless communication within the classroom using differentiated instruction. Indeed, this cohort's learning has been fashioned around their own specific skills and "talents" (Sparks, 2015). Differentiated instruction gave Millennials personalized and customized education from the outset of their academic trajectories. They learn best through computer screens, icons, sounds, virtual reality games, and show non-linear learning behaviors (Tapscott, 2009). The practice of differentiated instruction and the advancement of the Internet allows for the abovementioned customization tailored to their individual skills and preferences with one exception – standardized testing.

The context in which students study for exams can influence recollection during the target assessment. The study environment may adversely impact performance when the conditions differ from those in which the assessment is administered. Context-dependent memory is the enhanced level of recall when the state or "context" that is present at learning is also there at the time of retrieval or usage (Baddeley, Eysenck, & Anderson, 2009). Context includes time of day, physical state, as well as sensory stimuli like sounds and smells. One reason that ambient clues like sound and smell can be helpful to test takers is the "transferability" (Parker & Gellatly, 1997). The student may associate the song they listened to or the smell of the room while studying with a learned concept. Therefore, if the studying context is congruent with the test-taking context, students may be better able to demonstrate knowledge of the learned content while taking the test.

Millennials and Testing Preferences

Testing preference research exists in marketing, psychology, and economics, as three examples (Lichtenstein & Solvic, 2006). In education, research has tended to focus on assessment type and instructional delivery preferences. Assessment type preference can represent a student's "favored" kind of assessment question (e.g., multiple choice, essay). Instructional delivery preference refers to how the lesson is conveyed, meaning online versus in person or lecture versus small group instruction. A study found that university students preferred instructional method is a direct instruction lecture (Hativa & Bienbaum, 2000). The research is inconclusive as to students preferred assessment methods. One study has found that college students prefer traditional written assessments with closed questions compared to alternative assessments such as oral examinations and open-ended items (Van de Watering, Gijbels, Dochy, & Van de Rijt, 2008). A contrasting study states that many students view traditional assesment as capricious (Struyven, Dochy, & Janssesn, 2005). Evidence suggests that students' academic outcomes are improved when both their assessment type and instructional delivery preferences are considered (Dancer & Kamvounias, 2005; Wang, 2004). Building on this knowledge several measures of academic preference are used to measure student preferences.

Two preference assessments designed to measure college student learning preferences are the Learning Environment Preference and the Environmental Preference Assessment. The Learning Environment Preference measure assesses five domains – course content, instructor, peers, classroom activities, and course evaluation (Moore, 1989). The Environmental Preference Assessment is designed to assess preferred environments and how these contexts impact interpersonal communication amongst individuals (Richmond & McCroskey, 1995). Researchers have for years studied classroom environments and instructional settings, but few studies exist that have reliably measured students' preferred conditions for testing.

Organization of Part I: Pilot Study

The organization of Part I (Pilot Study) includes a Literature Review (Chapter II) of test environments, test performance factors, Millennial learning, multitasking, and differentiated instruction. The Methodology (Chapter III) contains a quantitative phase (Phase I) examining the psychometric properties of the Test Environment for Optimal Performance (TEOP) measure in a college student sample, and a qualitative phase (Phase II) exploring the perceptions of standardized testing in a sample of recent high school graduates. Chapter III is organized in the following manner: (1) Pilot Study Phase I (Quantitative) including the purpose and research questions, participants, measures, procedure, and data analyses, and (2) Pilot Study Phase II (Qualitative), containing the purpose and research questions, approach, participants, data collection, data analysis, trustworthiness, and ethics. Following the Methodology, the Results (Chapter IV) for the Pilot Study Phase I (Quantitative) will be reported, followed by the Pilot Study Phase II (Qualitative) findings. Finally, in the Discussion (Chapter V), the Pilot Study Phase I (Quantitative) and Pilot Study Phase II (Qualitative) results will be discussed, which includes implications, limitations, and future directions.

CHAPTER II: LITERATURE REVIEW (PILOT STUDY)

This chapter reviews the body of literature exploring Millennial college students' test environment preferences. In this Literature Review, topics covered include Millennial learning, multitasking, differentiated instruction, test performance factors, time of test, as well as context-dependent memory consisting of study/test environment incongruences. Research relating to environmental influences on cognition such as music/sound influences on cognition, movement/action preferences, and existing or similar measures of testing preferences are also detailed.

Millennial Learning

The Millennial generation, current college students, are vastly different than their predecessors. Born between the years 1981 and 1996, these students exhibit traits and preferences that can be linked to their unique upbringing (Dimcock, 2018; Howe & Strauss, 2009). Millennials are inundated with material objects. They grew up in homes that had 50% more "things" (measured in weight) than houses did in 1980 (Howe & Strauss, 2009). They have developed as somewhat different students than young people of past generations due to political climate, societal changes, and family life. Veen and Vrakking (2006) proposed the term "Homo Zappiens" to label this new generation of learners who learn in a considerably different way from their predecessors. This generation naturally developed the cognitive skills necessary for learning (e.g., enquiry-based, discovery-based, networked, experiential, collaborative, active, and self-regulative) without formal instruction (Veen & Vrakking, 2006). The above described cultural phenomena as Millennials came of age have made them greatly dissimilar from preceding generations.

Of the many differences Millennials exhibit from previous generations, student learning preferences are among the most pronounced (Howe & Strauss, 2009). Research suggests this generation favors a variety of active learning methods. When Millennials are not engrossed in something, their attention quickly shifts elsewhere. They have grown up with a huge assortment of choices and consider these options their birthright. Millennials are a generation of impatient, experiential learners, digital natives, multitaskers, and gamers who expect constant roaming connectivity (Sweeney, 2006). Not only do they appreciate the opportunity to make choices (Kirschner & Van Merrienboer, 2013; Sweeney, 2006), they demand ultimate consumer control, which is personalization and customization to meet their needs, interests, and tastes.

Customization means giving students control over certain facets of their learning environment, which is expected to produce favorable effects on learner motivation, and in turn, may increase learning outcomes (Wolters, 2003; Zimmerman, 2002). Research involving the generations before Millennials has evidenced that learner customization results in increased academic performance (Kinzie, Sullivan, & Berdel, 1992). More recently, Kalyuga and Sweller (2005) conducted a study on support given for subsequent tasks adapted to the learner. One group had learner-adapted Algebra instruction to develop expertise using formative quizzes. The control group received a standard Algebra instructional method. The personalized learning group showed gains in Algebra skills from pretest to posttest and higher cognitive efficiency compared to the control group (Kayluga & Sweller, 2005). This evidence indicates that personalization of instruction leads to positive academic outcomes. Thus, the lifelong exposure to instantaneous customization has formed the educational behaviors and preferences of the Millennial generation.

Multitasking

Prior to the mid-1980's, whether it was printed word or on a computer screen, information was presented linearly, one topic at a time. The invention of operating systems such as Windows allowed for the simultaneous availability of multiple documents and tasks (Beaton, 2017). Students today from Kindergarten through college represent the first generation to grow up with this operating system technology. Millennials have spent their entire lives using computers, videogames, digital music players, cell phones, and other digital tools. This group has been raised in a society inundated with multimedia and has developed into multitaskers (Price, 2009). Multitasking is defined as attempting to do multiple things simultaneously, or to switch from one task to another in immediate succession (American Psychological Association, 2006). These multitasking skills allow Millennials to function differently than previous generations (McGlynn, 2005). This group is unique in how they perform required tasks; rather than focusing on one project at a time, Millennials work and study while juggling social and entertainment options.

The current generation of young adults has an advanced relationship with technology that was formed at birth (Beastall, 2008). Even toddlers are developing multitasking strategies via technological familiarity that enable them to navigate novel environments (Presenky, 2001). Using technology with regularity allows young children to experience how sounds, images, and texts interact, and may be crucial to early schooling success and overall development in the modern, digital world (Presenky, 2001). For older students, the smartphone is ubiquitous on college campuses and is commonly used in educational settings (Lepp, Barkley, & Karpinski, 2014). Studies have revealed that about two-thirds of university students (*N* = 1,026) reported using electronic media while in class, studying, or doing homework (Jacobsen, & Forste, 2011). Despite these mutually exclusive activities, Millennials use smartphones in the classroom as both a leisure-time and learning tool simultaneously.

Millennials prefer a less formal educational environment (Price, 2009) and excel at juggling several tasks at once (Sweeney, 2006). They believe multitasking can accelerate their learning by allowing them to undertake more than one task at the same time. For example, a student may download and listen to a lecture while doing his/her laundry or exercising, and Millennials will almost never instant message someone without doing some other task(s) simultaneously (Sweeney, 2006). Educators in pre-Millennial cohorts do not believe their students can learn successfully while watching TV or listening to music, because they cannot learn under those conditions (Presenky, 2001). That is, pre-Millennials have not practiced multitasking constantly during their formative years as Millennials have done.

Kirschner and Van Merrienboer (2013) argue that the human brain is unable to multitask, and what is referred to as multitasking is actually task switching. Instead of true multitasking, what Millennials actually do is pause one task in order to undertake another and then switch back. This switch requires attention and forces a person to shift mental processes from accessing one mental schema to another (Kirschner & Van Merrienboer, 2013). Via extensive multitasking, Millennials are rewiring their brains to reduce performance deterioration by increasing the speed at which the brain processes information. This rewiring enables tasks to be processed more rapidly (Copper, Seppanen, & Gualiteri, 2012), and provides some evidence in support of this generation's adept multitasking.

Differentiated Instruction

Multitasking Millennials prefer academic customization and individualized instructional methods based on their abilities as learners. Differentiated instruction and assessment, or differentiation, provides individual students with varied avenues to learn (i.e., often in the same classroom). These different paths include how students acquire content, process ideas, and develop teaching materials and assessment measures. The goal of differentiation is for all students within a classroom to learn effectively, regardless of differences in ability or personal preferences (Tomlinson, 2001). Differentiated instruction allows all students to access the same curriculum by providing entry points, learning tasks, and outcomes that are tailored to students' unique needs (Subban, 2006).

This type of instruction acknowledges that learners can vary by culture, socioeconomic status, gender, motivation, ability/disability, personal interests, and more (Nunley, 2006). By considering unique student learning needs and diverse characteristics, teachers develop targeted and personalized instruction for effective classroom learning (Tomlinson, 1999). An analogy to describe differentiated instruction is the “shuttle bus,” which collects riders at their individual locations and transports the group to their final destination. Similarly, differentiated instruction meets each learner at their specific level with the same endpoint for all students in the classroom.

Since the 1990's, differentiated instruction has become a common technique applied to diverse students learning together (Sparks, 2015). As U.S. schools have become more varied with the introduction of diversity and disability inclusion, differentiated instruction has been implemented at all grade levels for students at all ability levels (Hatfield, 2017). Differentiated instructional models have evolved with technological advances, making it easier to develop and monitor education plans for many students at once. Teachers are expected to design education plans that enhance and complement self-regulated learning, with student choice as a central component in these designs (VanHout-Wolters, Somons, & Volet, 2000). Through constantly offering options, Millennials have been conditioned to expect a level of cusomization that has been consistently present over the course of their educational histories.

A student’s learning style refers to the favored way in which a person absorbs, processes, comprehends, and retains information. Differentiated instruction recognizes that students have their own personal learning style. Individual learning styles are influenced by cognitive, emotional, and environmental factors and/or preferences, as well as life experiences. Additionally, some preferences include direct instruction, practice work, movement, interaction with peers, and verbal discussions (Kise, 2007). Research has noted that personalization of tasks yields more efficient and effective learning than a fixed sequence of that is identical for all learners (Corbalan, Kester, & Merrenboer, 2006). In addition, student control over their own learning is believed to make the process more appealing.

The area of differentiated instruction is also applicable to personlized assessment. In the past, assessment was seen primarily as a way to assign grades or to uncover if students had met intended objectives. Over time, assessment has become moreso a vehicle for learning (Dochy & McDowell, 1997). The current trends blend instruction and assessment (Segers, Dochy, & Cascallar, 2003) where teachers and students might work together to develop appropriate assessment methods to fit the individual learner. Overall, personalized instruction and assessment, now commonplace in U.S. schools, have the potential to decrease test performance factors, both overt and covert, that are barriers to academic success.

Test Performance Factors

There are extrinsic influences that create obstacles to measuring the real performance of the student. Assessment allows for measurement of ability, but the test takers may be affected psychosomatically by circumstantial factors (i.e., environmental conditions) impacting the reliability and validity of test scores (Zhu & Han, 2011). For example, just as individuals have a physical preference for writing with either their right or left hand, individuals have psychological preferences for instruction and testing. Human beings have psychological preferences for increasing energy, taking in information, and decision making; all of these preferences process in education (Kise, 2011). Extrinsic influences that impact true performance include thirst, chewing gum, desk type, movement, anxiety, and sounds. These psychological preferences and external influences may increase measurement error if incongruent with test conditions.

Research has supported that bringing water to an examination room is linked to improved student grades (Pawson, Doherty, Martin, Soares, & Edmonds, 2012). The researchers concluded that accessing and drinking water may have physiological benefits for cognitive processing that lead to reduced test anxiety and enhanced performance (Pawson, Doherty, Martin, Soares, & Edmonds, 2012). Conversely, Ritz and Berrut (2005) found that dehydration significantly reduces attention, concentration, and short-term memory (Ritz & Berrut, 2005). Thirsty test takers have increased drowsiness and headaches, which also prolongs their response time to exam questions. These findings suggest that concentration and memory performance can be improved by drinking water.

Evidence suggests that children who drink water during class achieve better results on visual attention tasks. Edmonds and Jeffes (2009) found that when given additional water beyond their normal intake, children perform better on visual attention tests (Edmonds & Jeffes, 2009). Children were tested both after drinking 168ml of water and without water. The tasks included a letter cancellation task, ball catching, and a game using the Nintendo Wii console. Children had significantly higher scores on the computer game, as well as better scores on the letter cancellation task after drinking water. Bringing water into a test is associated with an increase in the students' grades. Students with water scored an average of five percent higher ($N = 477$) than those without. After controlling for aptitude, a ten percent score improvement could be predicted for water drinkers (Pawson, Doherty, Martin, Soares, & Edmonds, 2012). Furthermore, consuming more than 200ml of water improved ball catching skills (Booth, Taylor, & Edmonds, 2012). This research provides evidence that drinking water offers benefits that improve performance on both visual attention and fine motor skills.

Similar to the above, a study from the Baylor College of Medicine indicated that chewing gum has a positive impact on teen academic performance. Students who chewed gum had a statistically significant three percentage point improvement ($N = 108$) on math standardized test performance and better final grades than those who did not chew gum (Johnston, Tyler, Stansberrry, Palcic, & Foreyt, 2009). Other studies have found that chewing gum may help individuals concentrate for longer periods of time (Morgan, Johnson, & Miles, 2013). The results showed that those who chewed gum had quicker and more precise reaction times than those who did not chew gum. This was especially evident at the end of a task (Morgan, Johnson, & Miles, 2013). That is, at the end of a test, those students who were chewing gum continued to improve and maintained their performance level compared to those who did not chew gum. (Tanzer, von Fintel, & Eikerman, 2009). Overall, the evidence indicates that chewing gum aids academic performance and other cognitive abilities.

Research has indicated that there are neurocognitive benefits of stand-height desks in classrooms, when students are given the choice to stand at their desk or sit based on their individual preferences. The results found that use of standing desks was associated with significant improvements, between 7%-14% ($N = 41$) in executive function (e.g., organization, focus, self-monitoring) and working memory (Mehta, Shortz, & Benden, 2015). Additionally, standing school desks are also associated with enhanced student focus in elementary school children (Koepp, Snedden, Flynn, Huntsman, &

Levine, 2012). Other related qualitative studies show that teachers report improved student attention when using standing desks (Mehta, Shortz, & Benden, 2015). Overall, the collection of findings supports an academic benefit of standing desks in school settings.

Providing options for student movement while learning has increasingly been used in classrooms. Movement while learning optimizes focus and attention, allowing a release of energy so that concentration is directed toward the assessment at hand. Exercise and/or physical movement has been linked to improved cognitive functioning in children (Wechsler, Devereaux, Davis, & Collins, 2000). Similarly, research in both animal and human studies, has shown that exercise enhances cognition and brain functioning (Tomporowski, Davis, Miller, & Naglieri, 2008). Amplified local cerebral blood flow associated with exercise has been shown to improve cognitive functioning and working memory (Perreira, Green, Nandi, & Aziz, 2007). Additionally a recent study found that walking improves creative thinking by up to 81% ($N = 48$) compared to sitting (Oppezzo & Schwartz, 2014). Adults have the autonomy to move readily available. They can fidget, doodle, stretch, or get up and walk around (Hess, 2015). Overall, a majority of classroom instructors have understood the value of student movement for years, but freedom of movement is generally unavailable to student standardized test takers.

Test anxiety is an important factor impacting standardized test outcomes as it has debilitating effects on performance (Hill & Wigfield, 1984). This innate trait (Spielberger & Vagg, 1995) is the condition of extreme stress, nervousness, and discomfort during and/or before taking a test. This apprehension creates significant barriers to learning and academic performance (Andrews & Wilding, 2004). Numerous studies have shown that children in all demographic subgroups can be impacted negatively by test anxiety (Hill & Wigfield, 1984; McDonald, 2001). A meta-analysis of 562 studies has shown that highly test-anxious students score approximately 12 percentile points less than their low-anxiety peers (Hembree, 1988, Salend, 2012; Vaez & Laflamme, 2008). Conversely, the mitigation of test anxiety can lead to improved academic outcomes.

Another test performance factor is listening to music, as this is related to reduced anxiety in highly stressful situations (Rastogi & Silver, 2014). Using smart tablets, MP3 players, and smartphones, students oftentimes listen to music while studying (Vogel, Verschure, van der Ploeg, Burg, & Raat, 2009). Research has shown that studying and testing in the presence of background music, especially music without lyrics, can reduce anxiety, improve performance, and increase concentration (Sheer, 2011). Music's association with decreased anxiety and enhanced attentiveness suggests it may mitigate negative test performance factors and result in improved outcomes.

Listening to music activates the memory centers in the brain, such as the hippocampus and parts of the frontal lobe, and is associated with an improved ability for auditory imagery (Passion, 2017). The powerful link between music and memory is related to large areas of the brain activated by music. Brain imaging has shown that music activates the auditory, motor and limbic (emotional) regions of the mind (Alluri et al., 2013). A Florida International University study found that some children with ADHD had increased success when listening to music while they did homework (Pelham et al., 2011). Additionally, music may improve surgeons' task focus. For example, music is played between 62% and 72% of the time in operating rooms during surgical procedures, with over 75% of the operating room staff ($N = 100$) believing that music reduces anxiety and improves efficiency (George, Ahmed, Mammen, & John, 2011, Gregorie, 2014). Overall, the evidence is in favor of the positive impact of music on memory and performance for both children and adults.

Research on the impact of distractions on learning suggests the ability to concentrate on the task at hand can be diminished by outside diversions (Gazzaley & Rosen, 2016). Unsurprisingly children in visually stimulating environments spend more time off task during class time and made less academic gains than those in rooms with fewer optical distractions (Fisher, Godwin, & Seltman, 2014). Further evidence suggests that listening to music while reading results in more rereading and interrupted processing (Zhang, Miller, Cleveland, & Cortina, 2018). Although, studies have presented some evidence that there is some academic benefit related to "rereading," this is independent of background music. Overall, the research provides evidence that minimizing distractions when studying can be beneficial to the learner.

Time of Test

Another component of academic performance is clock time as it relates to the natural sleep cycles of teenagers. "Time of test" refers to the time of day that standardized tests are proctored. Both the PSAT and ACT are administered in the morning with stringent protocols that the three-hour and 30-minute test must be completed before lunch (ACT, 2017, The College Board, 2017). Research suggests that during adolescence biological changes dictate both sleep duration (i.e., of approximately nine hours) and later wake and bedtimes (Crowley et al., 2014). While most children and teens need nine hours of sleep throughout adolescence, data show that older adolescents are biologically inclined to stay awake later than younger ones (Crowley et al., 2014). This tendency, in turn, makes them "groggier" during the same hours standardized tests are administered.

Additional research has recommended that schools begin at ten in the morning (i.e., at the earliest) to allow the academic schedule to better align with biological wake-up times of a typical teenager (Kelley, Lockley, Foster, & Kelley, 2015). Similarly, the American Academy of Pediatrics endorses later start times for middle and high schools to allow optimal levels of sleep and improved academic performance (American Academy of Pediatrics, 2014). However, as noted previously, the ACT requires all test takers arrive by 8:00 in the morning with testing beginning at 8:30 (ACT, 2017). Two related studies found improved academic scores in the afternoon compared to the morning. A study of high school seniors indicated that the students performed better in the afternoon than in the morning on tests and logical reasoning tasks (Hansen, Janssen, Schiff, & Phyllis, 2005). The second study, encompassing multiple grade levels, discovered that morning grades and test score performances were lower compared to afternoon classes (Cortes, Bricker, & Rohlfs, 2010). Overall, the evidence consistently demonstrates that later school start times are beneficial for improved school performance in adolescence.

Context-Dependent Memory

Decades of research on context-dependent memory allow for the inference of an association with preferred testing environment. Context-dependent memory is the improved recall of specific episodes of information when the context present at encoding and retrieval are the same (Tulving & Thompson, 1973). That is, environmental context-dependency effects occur when there is improved performance on a memory-type test when it occurs in the same environment in which the tested material was initially learned (i.e., compared to the test occurring in a different environment; e.g., matching condition versus mismatching condition; Grant et al., 1998). Context can be internal (i.e., a state of mind at the time of memory encoding), temporal (i.e., the time of day when a memory was encoded), or external (i.e., the physical surroundings).

Numerous studies have found that information recall is improved when the context is matched at encoding and retrieval (Grant et al., 1998; Godden & Baddeley, 1975; Smith & Vela, 2001; Tulving & Thompson, 1973). For example, in one investigation, participants were asked to study under either quiet or noisy conditions. Afterwards, they were asked short-answer and multiple-choice questions on the previously learned material, which prompted both recognition and recall. The participants whose noise levels were matched during studying and testing conditions remembered significantly more information than the mismatched noise levels (Grant et al., 1998). In a related study, Smith (1985) investigated the effect of background music on context-dependent memory. Participants learned a group of words under one of three background music conditions: a classical piece, a jazz selection, or a quiet condition. Following the initial learning of the word list, contributors completed a memory test in one of two conditions: same context or different context. Smith found that participants recalled significantly more words when the same musical context was reinstated during testing (Smith, 1985). The results indicated that context-dependent memory related to background sound results from contextual cuing rather than a damaging effect caused by the distractions of background noise during testing.

In another investigation of external context-dependent memory, scuba divers learned word lists on land or underwater. Their memory for these words was tested afterwards in either the same or the opposite context. It was found that those divers who had to recall the words in the original environment remembered significantly more words than those who changed conditions (Godden & Baddeley, 1975). This suggests that the context present at conditioning and the ability to demonstrate recollection is closely linked to the context at time of desired recall.

Related to the above literature on external contexts, internal contexts (i.e., "states") such as stress can also impact learning and retrieval. Across studies, stressed participants perform worse than non-stressed controls when learning and testing contexts are dissimilar. However, if the learning and testing environments are matched, the negative effect of stress disappears (Schawbe & Wolf, 2009). Typically, individuals show improved memory performance when the learning environment or "state" is restored at testing compared to when assessment occurs under different environmental conditions (Smith & Vela, 2001; Tulving & Thompson, 1973).

Measurement of Preferences

Related to the above literature is the measurement of a learner's preferences. A preference is the greater liking of one option over an alternative. The "acceptability" or fondness of one product/condition over another is the primary goal in preference measurement. Preference testing often poses questions such as, “Which do you prefer?” Two main types of preference testing include a Simple Preference Test or a Ranking Preference Test. In a Simple Preference Test, the researcher provides options and asks subjects, “Which do you prefer?” A Likert scale can also be used in a Simple Preference Test, for example, by including a range from “Always” to “Never.” This response scale addresses how frequently an individual prefers a provided option, and incorporates more detail compared to a dichotomous response-option of “Yes” or “No.” A Ranking Preference Tests asks participants to put three or more options in order of preference (Walkers, 2004). Ranking Preference Tests can be used to measure students’ favored assessment types. Van de Watering and colleagues (2008) note that preference indicates choice, and in most cases, students cannot choose assessment formats. Due to the lack of choice, it is surmised that preference research likely only measures what assessment format the student believes is the most suitable design to measure their specific ability and not a true preference (van de Watering, Gijbels, Dochy, & van der Rijt, 2008). While assessment preference rankings are often theoretical, these perceptions may exert influence on individual effort and test scores.

Students’ perspectives on assessment tasks can impact motivation and learning outcomes (Alkharusi, 2013). That is, the way one thinks about learning influences how a student approaches assessment responsibilities (Struyven, Dochy, & Janssesn, 2005). An empirical study of 264 first-year law students found that improved academic performance resulted when students were allowed to be involved in the assessment process (Dancer &

Kamvounias, 2005). Additionally, findings from a study of 503 English as a Second Language students indicated that pupil perception of a learning-oriented assessment environment contributes positively to meeting performance goals. Conversely, perceptions of test-oriented assessment environments negatively impact performance goal outcomes (Wang, 2004). These studies suggest that the perception of choice can result in improved assessment outcomes for students.

Students commonly have negative perceptions of traditional or “normal” assessment methods. Many have the opinion that “normal” assessment methods have a severe negative impact on learning (Sambell, McDowell, & Brown, 1997). In contrast, these same students viewed alternative assessments as enhancing the learning process (Sambell, McDowell, & Brown, 1997). Many students view traditional assessment as irrelevant and arbitrary, leading them to only learn for the purpose of the specific assessment with no plan of retaining the knowledge in the future (Struyven, Dochy, & Janssesn, 2005). The abovementioned literature suggests that many students have developed their own unique, strongly held perceptions and beliefs on assessment formats. These individual preferences provide for distinctive approaches to learning and testing.

CHAPTER III: PILOT STUDY METHODOLOGY

The pilot study implemented a two-phase process. The first phase (Phase I) involved the creation and dissemination of a measure entitled the Test Environment for Optimal Performance (TEOP) to current college students. The data were analyzed using single-sample cross-validation (i.e., an Exploratory Factor Analysis [EFA] followed by a Confirmatory Factor Analysis [CFA] with a single, split sample) to determine the psychometric properties of the TEOP. The second phase (Phase II) of the Pilot Study used qualitative methods and analyses in a basic interpretive design. Recent high school graduates participated in this second phase and results provided more evidence in support of the findings from the first phase.

Chapter III is organized in the following order: (1) Pilot Study Phase I (Quantitative) containing the purpose and research questions, participants, procedures, measures, and data analysis, and (2) Pilot Study Phase II (Qualitative) comprising the purpose and research questions, approach, participants, data collection, data analysis, trustworthiness, and ethics. The research question guiding the Pilot Study Phase I (Quantitative) is, "What are the psychometric properties (i.e., content and construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a university student population?" The main research question guiding the Pilot Study Phase II (Qualitative) is, "What are recent high school graduates' perceptions of their experiences with current standardized testing practices and environments?"

Purpose and Research Question: Pilot Study Phase I (Quantitative)

With the proliferation of technology and social media, society, and education have transformed substantially in the past decade (Oblinger, 2008); however, the rules and procedures for proctoring and taking exams remains unchanged (ACT, 2017; The College Board, 2016). Additionally, the emphasis placed on standardized testing for both students and schools has never been greater (Meador, 2017). Thus, it is important to construct measures with valid and reliable scores of self-reported optimal performance conditions. Development of these measures can give stakeholders a method with which to assess current students' (i.e., Millennials') testing preferences when interpreting high-stakes standardized assessment scores.

The purpose of Pilot Study Phase I (Quantitative) included investigating the psychometric properties of a newly-developed measure entitled the Test Environment for Optimal Performance (TEOP) in a sample of university students. Through cross-validation methods and statistical analyses, validity, and reliability of the scores on the TEOP were obtained. The main research question in Pilot Study Phase I (Quantitative) was, "What are the psychometric properties (i.e., content and construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a university student population?"

Content validity is the degree the items within the measure sufficiently represent the construct being assessed (Crocker & Algina, 2008). Content validity evidence includes relevance, representativeness, and technical quality (Messick, 1995) Construct validity is the extent a test measures the theoretical concept it is intended to measure

(Brown, 2009). Convergent and divergent validity are the two categories of construct validity (Dimitrov, 2010). Convergent validity is the degree that similar items that should be related are related. Divergent validity is the degree that unlike items that should not be related are not related (Messick, 1995). Because a construct is theoretical, construct validity can never be fully proven (Cronbach & Meehl, 1955). Internal consistency reliability is a score based on correlations of different items on the same test that measures whether items that purport to assess the same construct produce comparable results (Cronbach L. , 1951).

A measure was constructed originally containing 11 items gauging the preferred environmental conditions for optimal performance while testing. The item response format was a 5-point Likert scale rating students' personal preference of a particular environmental condition while taking a test anchored by "Never Me" and "Always Me." The TEOP was administered with directions to use the abovementioned Likert scale to indicate the degree to which each statement represents the respondent's test taking preferences.

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were used to analyze the data in this first phase of the Pilot Study. The following paragraphs include descriptions of the context, participants, procedures, data, and analyses that were used to address the abovementioned research question.

Context: Pilot Study Phase I (Quantitative)

The population in the current study straddles two generations: (1) "Millenials" and (2) "Generation Z." Pew Research Center defines Millennials as those born between

1981 and 1996. Anyone born in 1997 and beyond is categorized as Generation Z (Dimcock, 2018). Both groups are different from their predecessors and exhibit traits and preferences that can be connected to their distinctive upbringing (Howe & Strauss, 2009). Millennials have been raised from birth using technology, specifically computers, videogames, digital music players, cell phones, and other multimedia tools. For example, college students have spent less than 5,000 hours reading, but over 10,000 hours playing video games, and over 20,000 hours watching the television (Prensky, 2001). However, compared to newer generations, Millenials used dial-up modems as children and had to wait patiently for several minutes to use the Internet.

The younger Generation Z students are more likely to be task-switchers than previous generations (Kirschner & Van Merrienboer, 2013). They do school work with multiple distractions in the background and seem to easily switch from work to play. Members of Generation Z often study with videos or music on while simultaneously “face-timing” friends (Beall, 2017). This cohort contains active learners who expect to be fully immersed in the learning process (Barnes and Noble College, 2015). Overall, both generations have come to expect constant access to technology and demand individualized services in both work and play (Prensky, 2001; Tapscott, 2009).

The abovementioned generations, who have only lived in a computer technology-infused world, are both generally viewed as “multitaskers” (Price, 2009). As technology is a natural part of their environments, these cohorts of “multitaskers” have perpetual connectivity (Oblinger, 2008). In accordance with the barrage of multimedia and incessant multitasking, these college students are accustomed to an incalculable assortment of options. They have grown to expect other people and/or institutions to give them more flexibility, personalization, and customization to meet their individual needs (Sweeney, 2006). Despite the generational preference for individual choice, the examination procedures used in large-scale, high-stakes standardized testing conditions (e.g., the ACT) have been stagnant for decades (ACT, 2017).

Testing, as a fundamental part of the education system, is defined as the process for assessing a sample of an individual's performance as demonstrated by an aptitude test of optimal performance or an inventory test of typical performance (Crocker & Algina, 2008). For example, the ACT, introduced in 1959, is a standardized test used to measure college readiness in the United States (US). The ACT website expressly states that is not an aptitude or IQ test but rather items are directly related to material taught in high school courses (ACT, 2017). At the same time the ACT is used by colleges and universities to make decisons on an indvidual's ability to succeed in undergraduate education, because of this for puroposes of this study ACT scores will be referred to as a measure of aptitude.

The testing procedures and enviromental rules for taking the ACT have not varied since its inception (ACT, 2017). The forthcoming generations of test takers have been conditioned to expect customization and individual choice in all areas of life. These entitlements are contrasted with testing companies that in an effort to minimize random error provide little variety in test environment conditions.

There are extrinsic influences that create obstacles to measuring the actual performance of the student during a test. These factors include temperature, light, sound inside/outside the examination room, seating arrangements, and the testing staff (Rasul & Bukhsh, 2011). When students are impacted physically or psychologically by peripheral stimuli such as the environmental conditions noted above, the validity and reliability of students' test scores (i.e., their actual abilities) are tenuous (Crocker & Algina, 2008; Zhu & Han, 2011). Scores may be affected detrimentally for students whose preferred testing environment is different from the customary exam specifications. Conversely, students whose preferred testing environment is congruent with traditional high-stakes assessment settings may receive an inadvertent advantage over others. These "one-size-fits-all" testing practices presume that all students prefer and can benefit from uniform assessment conditions (Johnson, 2008).

Classical Test Theory postulates that all people have a true score that would be achieved if there was no measurement error (Allen & Yen, 2002). All observed test scores are a composite of two hypothetical elements, a true score and an error component. Tests contain both systematic and random error. Systematic errors are those which regularly affect an individual's score because of some particular characteristic of the person or the test that has nothing to do with the construct being measured (i.e., A scale is not calibrated properly). Systematic errors reduce the accuracy and utility of the test but are consistent and repeatable (Allen & Yen, 2002). Random errors affect an individual's score in a positive or negative direction because of pure chance happenings (i.e., guessing, distractions, administration errors, fluctuation of the individual examinee's state). Random errors reduce both the consistency and the usefulness of the test scores (Allen & Yen, 2002).

Testing companies by establishing a uniform assessment environment are able to minimize but not eliminate random error. The standard testing conditions may increase systematic error for a student who is more comfortable under different conditions. Reducing random error over systematic error is emphasized because systematic error is more consistent than random error and easier to address (Taylor, 1999).

This study's population of interest included current college students, who are categorically Millennials and the forthcoming Generation Z, who were raised in a time of tremendous technological change. In the current study, these populations, although from varying backgrounds and assorted geographic regions in the US, were predominantly from the Midwest. Beginning with the 1929 Middletown study, researchers and social commentators have used the Midwest as a "typical" representation of the US (Scheetz, 2000) It is possible that the changes in technology might also carry over to assessment condition preferences and the unchanged test format may result in systematic errors for test takers based on individual traits not present in prior generations.

Participants and Procedures: Pilot Study Phase I (Quantitative)

A total of 666 university students completed the TEOP. Forty-nine cases were removed due to incomplete or inconsistent information provided. The removal of these cases and description of the final analysis sample will be presented in detail in the Results section (Chapter IV) below.

The scale construction and administration procedures followed the recommended guidelines from Clark and Watson (1995), Allen and Yen (2002), and Crocker and Algina (2008). Details are provided in the Measure section below. First, the definition of the domain of interest (i.e., test environment preferences) was examined. This included a review of literature and examination of test center rules for major standardized assessments in the US. Next, items related to test environment preferences were created. A response continuum (i.e., "Never Me" to "Always Me") and the number of response choices (i.e., five) were selected for the Likert scale. Following Institutional Review Board (IRB) approval (Kent State University IRB # 16-752)", items were reviewed by a psychometrician for criteria (Fowler, 2002).

The measure was uploaded to a survey-hosting website addressing participant demographics and the 11 items of the TEOP. Undergraduate university students were recruited to participate via social media and email. A short recruitment message was used, "Your participation in this short survey study can help identify student preferences for optimal test performance, which can impact best practices for higher education and the workforce. Your responses will be completely anonymous, and no identifying information will be collected. The online survey was open for one month before it was closed at the end of the study.

Measure: Pilot Study Phase 1 (Quantitative)

The following paragraphs include the construct-based test construction of the Test Environment for Optimal Performance (TEOP). The TEOP is an exploratory measure that originally contained 11 items using a Likert response format to measure the preferred environmental conditions for optimal performance while testing. A principal objective of scale development is to create a valid measure of the construct (Clark & Watson, 1995), or an underlying concept that cannot be measured directly that is used to explain human behavior (Crocker & Algina, 2008). Clark and Watson (1995) proposed four steps in construct-based test construction: (1) Conceptualization, (2) Literature Review, (3) Creation of an Item Pool, and (4) Choice of Format.

Conceptualization is the establishment of the desired construct and its theoretical framework (Clark & Watson, 1995). The underlying construct contained in the TEOP is preferred testing environment. This construct was defined as preferences for testing conditions such as noise level, food and beverage availablity, and permission to move around during testing. Other dynamics such as behaviors and expectations of the primary demographic composition of current university students and the Millenial generation were considered as well. That is, university students' preferred environmental conditions for testing include different noise tolerances for background sounds, music, and televsion. These environmental condition preferences also involve varying physical behavior expectations including drinking water, eating food, chewing gum/candy, and other options for movement during a test.

Following the conceptualization of the theoretical framework of test environment preferences, a literature review of the construct and related concepts was conducted. A literatrure review should examine preceding methods to asses similar constructs and progress to broader concepts related to the topic of interest (Clark & Watson, 1995). The literatrure review focused on Millennials' behaviors, differentiated instruction, and the impact of water and gum chewing (and related actions) on concentration and anxiety relief during studying and/or testing. This literature review provided confirmation of the hypothesized theoretical framework for the creation of an item pool.

The third step in construct-based test construction is the creation of an item pool. The items were constructed to include all likely elements that might involve the construct of interest (Loevinger, 1954). In addition, several item-writing rules were followed. The items were written in plain language so as to be understood by the intended participants (Crocker & Algina, 2008). In addition, the items permitted variability in responses such that one response is not selected by everyone (Clark & Watson, 1995; Crocker & Algina, 2008). "Double Barrelled" items that address more than one concept were not used (Clark & Watson, 1995). Overall, the TEOP contained 11 plain-language items that comprised the exhaustive domain of prefered testing environment.

Likert response formats contain three or more answer options, and according to Comrey (1988), are more reliable, deliver more stable outcomes, and produce superior scales compared to dichotomous and other response formats options. Clark and Watson (1995) state that Likert scales do not necessarily enhance relaibility and validity by offering more response items (e.g., a 9-item scale compared to a 5-item scale). That is, reliability and valdity may in fact be reduced if the participants do not grasp the sometimes subtle distinctions in response options, which may unintentionally lead to random responding (Clark & Watson, 1995). The TEOP item response format used a 5-point Likert scale (i.e., "Never Me," "Rarely Me," "Sometimes Me," "Often Me," and "Always Me"), and participants rated how strongly a particular environmental testing condition was indicative of their personal preference.

Table 2

The Test Environment for Optimal Performance (TEOP) Items (N = 11)

#	Item
1.	When taking a test, I prefer silence.
2.	When taking a test, I prefer background noise/environmental sounds (e.g., sounds of the ocean, rain, birds chirping).
3.	When taking a test, I prefer to listen to music.
4.	When taking a test, I prefer to have a television on in the background.
5.	When taking a test, I prefer to have water/beverage available.
6.	When taking a test, I prefer to have food available.
7.	When taking a test, I prefer to chew gum or candy.
8.	When taking a test, I prefer to have the option stand up.
9.	When taking a test, I prefer to have the option to walk around.
10.	When taking a test, I prefer to use a computer.
11.	When taking a test, I prefer to use paper and pencil.

Data Analyses: Pilot Study Phase I (Quantitative)

The research question addressed for Pilot Study Phase I (Quantitative) was the following:

"What are the psychometric properties (i.e., content and construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a university student population?" To address this research question, data were analyzed using split-sample cross validation. Cross-validation involves randomly splitting the data into corresponding subsets in order to performing both exploratory and confirmatory measurement-related analyses (Cudeck & Browne, 1983).

In cross-validation, this method provides a process for testing the reliability of results obtained from the exploratory analysis with the results from the confirmatory analysis. As mentioned above, split-sample cross-validation involves randomly splitting the dataset and then comparing the resulting models (Pedhazur & Schmelkin, 1991). A model is generated typically using EFA with the first sample, and then the model is tested using CFA with the second sample (IBM, 2011). The value of cross-validation lies in the capability of this method to test how a set of parameter estimates obtained from one sample will reproduce the observed data from another (Bandalos, 1993). Cross-validation does not require the collection of two related but different samples from the same popualtion which can be time consuming and expensive.

As noted above, one data analytic strategy typically implemented with the split samples is Factor Analysis (FA). FA consists of two major classifications: (1) Exploratory Factor Analysis (EFA), and (2) Confirmatory Factor Analysis (CFA). EFA and CFA are both used to help comprehend the shared variance of observed variables that may be attributed to an unobserved construct (Dimitrov, 2010). EFA is a quantitative theory-generating data analytic strategy and no hypothesis for the underlying factor structure is needed (Crocker & Algina, 2008). CFA is a quantitative theory-testing data analytic strategy and a hypothesis for the underlying factors comprising the model and the items within each factor is required (Schumacker & Lomax, 2016). EFA was used to uncover the hypotheised theory that was measured by the CFA.

It is a common cross-validation technique to conduct an EFA to generate the hypothesied theory and follow up with a CFA from a split data-set (Chazdon, Allen, Horntvendt, & Scheffert, 2013). The combined use of EFA and CFA is reccommended to enhance construct-validity in the developement of new measures (Morgado, Meireles, Neves, Amaral, & Ferriera, 2017). In the current study, after removing random responses and outlier cases, the final analysis sample was randomly split using SPSS version 24 into two datasets to conduct EFA (i.e., in the first sample) followed by CFA (i.e., in the second sample). The data analytic strategies used for both EFA and CFA are detailed in the following sections.

Exploratory Factor Analysis (EFA)

There are five considerations that are included in a decision sequence provided by Thompson (2010) when conducting an EFA: (1) Selection of an association matrix, (2) Number of factors extracted, (3) Method of factor extraction, (4) Method of factor rotation, and (5) Computation of factor scores (if needed). The first four steps in the decision sequence will be reviewed, as the fifth step is not needed in the current study. The first consideration in the aforementioned sequence is choosing an association matrix type. Factor Analysis uses association matrices to compute relationships between data, and a decision must be made as to which matrix of associations (e.g., correlation, covariance) to analyze. The current study used the correlation matrix. Since the item response format was ordinal (i.e., Likert) and some of the data were nonnormal, a Spearman Correlation matrix was appropriate (Lehman, 2005). Spearman correlations evaluate the monotonic relationship between two ordinal variables (Spearman, 1910). Spearman correlations are based on the ranked values for each variable rather than the raw data (Myers & Well, 2003).

For the second EFA consideration, both eigenvalues and a scree plot were used to determine the number of factors to extract. The K1 Rule states that factors with eigenvalues greater than 1.00 should be retained (Guttman, 1954). Eigenvalues is a ranking of the total variance explained by the items that comprise a factor (Pett, Lackey,

& Sullivan, 2003). An eigenvalue of 1.00 indicates the explanation of the average variability of a single item. Therefore a factor with eigenvalue greater than 1.00 explains more variability than a single item (Nunnally & Bernstein, 1994). A visual review of the scree plot illustrates how many factors to retain prior to an “elbow” or sharp bend appearing in the graph (Cattell, 1966). Assessing both the eigenvalues and scree plot can assist in selecting the appropriate number of factors to extract. The eigenvalues and scree plot sometimes reveal different numbers of factors to extract. Costello and Osborne (2005) cite the scree plot as the best option for researchers to use. The scree plot comprises inspecting the graph of the eigenvalues and looking for the natural bend or elbow point in the data where the curve flattens out. The number of items above the “elbow” (i.e., not including the point at which the break occurs) is usually the number of factors to retain (Costello & Osborne, 2005). The literature suggests that if the number of factors to extract is unclear, more than one FA can be performed with varying numbers of extracted factors (Gorsuch, 1983, Tabachnick & Fiddell, 2001).

The third step in conducting an EFA is that an extraction method should be selected. There are a variety of extraction methods with Principal Components Analysis (PCA) and other FA options being the most common. PCA is the default extraction method in SPSS and other leading software programs. PCA extracts uncorrelated linear groupings of items and is inappropriate in identifying underlying latent constructs (Brown, 2009). This is because PCA does not parcel out the errors of measurement from shared variance and consequently overestimates the linear associations among variables (Pett, Lackey, & Sullivan, 2003). The other category of extraction methods uses the covariance between items to create either correlated or uncorrelated factors. This category of extraction methods contains Principal Axis Factoring (PAF), which was used in the current study. This method is suitable when the objective is to identify latent constructs or factors in a group of items (Brown, 2009). PAF is viewed as a superior extraction method when not all correlations are greater than .80. PAF analyzes common variance and does not assume perfect score reliability (Kline, 2013). Nunnally and Bernstein (1994) assert that a PAF solution will provide a more accurate estimate of the correlations (Nunnally & Bernstein, 1994).

The measurement level (e.g., nominal, ordinal, interval/ratio, etc.) of the items in relation to normality can assist in choosing the most appropriate extraction method. The items were Likert-scaled, ordinal and non-normal. Skewness and kurtosis are examined for these ordinal data to determine the degree of nonnormality. In the current study, normality statistics (i.e., the items were moderately nonnormal) established the selection of PAF as the appropriate extraction method with no distributional assumptions (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

The fourth consideration in the sequence necessitates choosing the appropriate rotation method. Rotation involves finding "a simple solution" (Bryant & Yarnold, 1995) that renders more interpretable factors (Brown, 2009). There are two main categories of rotation – orthogonal and oblique. Orthogonal assumes the factors are uncorrelated and oblique allows for correlated factors (Pett, Lackey, & Sullivan, 2003). As an assumption of uncorrelated factors is rarely met in social science research orthogonal rotation was not considered (Pedhazur & Schmelkin, 1991). As there was a hypothesized correlation between factors (if more than one existed), Direct Oblimin was selected as it is one of the most commonly used methods of oblique rotation and it is endorsed for relatively simple factor structures (Mueller & Kim, 1978).

Confirmatory Factor Analysis (CFA)

Compared to EFA, CFA is a theory-testing technique in which expectations are determined regarding the number of factors, the items that represent the given factors, and if the factors are correlated. CFA aims to establish if the initial theory (i.e., measurement model) is upheld (Thompson, 2010). CFA analyzes the degree to which the hypothesized structure of identified factors aligns with the data (Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991). Additionally, a CFA can assess the efficacy of the latent constructs identified in the EFA (Pett, Lackey, & Sullivan, 2003). As part of the current investigation (i.e., Pilot Study Phase I [Quantitative]), a CFA was conducted to verify the factor structure identified in the EFA and to provide further evidence of the construct validity of the measure.

CFA examines if the data confirm the hypothesized model (Schumacker & Lomax, 2016). In the current study, the hypothesized model is based on cross-validation procedures from the EFA using the split sample. That is, the EFA results define the first step of a five-stage CFA process – Model Specification. This first step is followed by Model Identification, Estimation, Testing, and Modification (Schumaker & Lomax, 2016). These steps, outlined by Schumacker and Lomax (2016), provide confirmatory evidence of the factor structure obtained in Pilot Study Phase I (Quantitative) and are detailed in the paragraphs below.

Model specification. Specifying the theoretical model, also referred to as the covariance structure, is conducted based on theory and previous inquiry (Schumaker & Lomax, 2016). This stage consists of specification of the number of factors within the data, the factors that are associated with the observed variables, the factors that are expected to correlate, the errors that are expected to correlate, and the factor loadings that should be held equal (Dimitrov, 2010). The goal is to define the best model that creates the sample covariance matrix *S*, and then to assess how closely the model fits the population covariance structure. When the sample covariance matrix *S* of the specified model is inconsistent with the population covariance matrix ∑, which is created from the population covariance structure, the model is considered misspecified (Schumaker & Lomax, 2016).

CFA models are frequently represented using path diagrams in which circles represent the latent variables (i.e., factors or constructs) and squares or rectangles represent observed variables. The single-headed arrows (i.e., →) are used to show a direction of assumed causal influence, and the double-headed arrows (i.e., ↔) are used to imply a covariance among latent variables. CFA can be represented using the following equation:

$$x = \lambda\xi + \delta \qquad [1]$$

where x is the vector of the observed variables *i*, λ (lambda) is the matrix of loadings connecting the latent variables ξ_i to the observed variables x_i, ξ is the vector of common factors, and δ is the vector that represents the measurement error (Montilla, 2004).

Model identification. Model identification establishes if the hypothesized model is identified. A model is considered identified when each parameter is distinguished (Thompson, 2010) where the degrees of freedom are equal to or greater than 1. The number of free parameters analyzed must be less than or equal to the number of distinct values in the matrix *S* in the model. Free parameters in the theorized model include factor loadings, measurement error covariances, and correlations among the constructs (Schumaker & Lomax, 2016). The number of distinct values in the matrix *S* can be determined using the following formula:

$$p(p+1)/2 \qquad [2]$$

in which p is the number of observed variables in the model. If the number of distinct variables in the sample matrix *S* is greater than or equal to the number of free parameters the model is considered identified (Schumaker & Lomax, 2016).

Model estimation. Model estimation is the step in which the researcher finds the suitable “fitting function” (Schumaker & Lomax, 2016) that aids to reduce the differences between the population covariance matrix ∑ and the sample covariance matrix *S*. When components in the sample matrix *S* minus the elements in the population matrix Σ equal zero ($S - \Sigma = 0$), then the χ^2 value will equal zero, which indicates the data has a perfect model fit. (Schumacker & Lomax, 2016). The Maximum Likelihood (ML) estimation method allows for statistical interpretation including significance testing and goodness-of-fit evaluation. ML estimation is recommended to produce factors that replicate the correlations in the population and is often the default setting in software programs (Thompson, 2010). This estimation method assumes that the data are normally distributed but can support some deviations from normality. In practice, the assumption of normality is frequently violated (Hu, Bentler, & Kano, 1992). However, research has shown that ML estimates of factor loadings are valid under a wide range of distributions (Anderson & Amemyia, 1988).

Model testing. Model testing defines how well the data fit the model and is examined through goodness-of-fit indices. Chi-Square is a frequently used goodness-of-fit index, which tests the residual amongst the covariance matrices of the sample and the model expected in the population (Dimitrov, 2010). The Chi-Square index alone, however, does not provide enough evidence to indicate satisfactory or unsatisfactory model fit (Bentler & Bonnett, 1980). Thus, other indices are considered. Among the other prominent goodness-of-fit indices are the Goodness-of-Fit Index (GFI) and the Adjusted Goodness-of-Fit Index (AGFI; Dimitrov, 2010). The Standardized Root-Mean Square Residual (SRMR) and the Root- Mean Square Error of Approximation (RMSEA; Steiger, 1990) are also frequently used indices to assess the model-data fit (Schumacker & Lomax, 2016). Together, an assortment of indices was used to calculate the fit between the data and the hypothesized model, produced from the prior EFA.

Model modification. Model modification is used when the model does not fit the data. Major changes to the model such as adding or deleting a path are only advocated in instances of non-significance or when there is previous research/theory to support the change (Schumaker & Lomax, 2016). Typically, an error covariance between observed variables can be used to provide better model-data fit. When a modification by way of adding an error covariance is made, justification from the literature should be included

(Schumaker & Lomax, 2016). After these modifications are made (i.e., one at a time), the model is tested again and re-run.

Purpose and Research Question: Pilot Study Phase II (Qualitative)

Following the completion of the quantitative phase of the pilot study, a second qualitative phase was completed to provide additional evidence relating to individual standardized testing preferences. The purpose of phase two of the pilot study was to investigate how recent high school graduates have experienced standard assessment conditions and what environmental testing conditions would provide for optimal testing performance. The following primary research question guided this part of the study, "What are recent high school graduates' perceptions of their experiences with current standardized testing practices and environments?"

Approach (Methodology): Pilot Study Phase II (Qualitative)

A basic interpretive qualitative design was chosen for this portion of the pilot study. In a basic interpretive design, the aim is to assess how the participants give meaning to a condition and discover their perspectives on the situation (Merriam, 2002). Meanings are constructed by people as they participate in the world (Crotty, 1998). The goal of a basic interpretive study is to gain knowledge of the intangible domain of the subjects in an effort to comprehend the meaning they give to the events being studied (Bogdan & Biklen, 1998). The data are analyzed to recognize patterns and themes that recur across the participants. The outcome is a descriptive account of the findings (Merriam, 2002). Interprative description is used to allow themes to develop from the patterns that appear collectivly across the participants, while at the same time asking broad questions of the data (Hunt, 2009).

This basic interpretive design was appropriate in this pilot study phase because the aim was to understand how recent high school graduates give meaning to their individual experiences of standardized testing environments. Marshall and Rossman (1995) recommend beginning qualitative inquiry by exploring their own personal experience with the phenomenon. Based on this, the investigation was initiated by the researcher's exploration of his own experiences with standardized assessment. This reflexivity is an effort to recognize any preconceived biases and beliefs. Through this reflexivity exercise, the investigator attempts to understand the essence of the experience without distorting the research with preconceived bias or experience.

Participants: Pilot Study Phase II (Qualitative)

This study focused on students' experiences with standardized testing environments, and purposeful sampling was used in the participant selection process. Purposeful sampling involves identifying and selecting individuals knowledgeable or experienced with the occurrence of interest, in this case standardized testing environments (Creswell & Plano Clark, 2011). Purposeful sampling is typically used for data collection in traditional qualitative studies to learn more about a specific group of individuals' experiences or points of view (Creswell, 2007). This sampling strategy was selected based on the research question and the purpose of the study.

Participant inclusion criteria for the study included the following: (1) Are at least 18 years of age, but not yet 21 years of age, at the time of the interview, (2) Have taken either the ACT, SAT, or both, (3) Have attended school in the US, and (4) Are willing to be audio recorded as part of the interview and documentation process. Four participants who met the inclusion criteria listed above were selected. Interviews were conducted in June of 2017 with two females and two males between the ages of 18 to 20 at the time of the interview. The four participants were all 2016 or 2017 high school graduates who were either currently attending college or started in the Fall of 2017. The four participants were from the same Midwestern state in the US.

Data Collection: Pilot Study Phase II (Qualitative)

The study was submitted to and approved by Kent State University Institutional Review Board (IRB #17-729). In compliance with the IRB and at the recommendation of Hatch (2002), prior to data collection (i.e., interviews), each participant was asked to sign a consent form. The consent form outlined the terms of the research process and stated that as a participant, they were free to leave the study at any time. Prior to the interview, participants were informed that the researcher would not use any identifying information about the contributor. That is, all names and other identifying data were coded to protect the participants' anonymity.

The participants were also informed that they may be contacted in the future for clarification or follow-up questions. All interviews took place at a location of the participants' choosing. Each audio recording was labeled with a number code and no identifying information. The recordings were locked in a private office, inside a workplace that requires a punch code, in a locked building with 24-hour armed security guard. The semi-structured interviews began after informed consent was explained and obtained.

The open-ended, semi-structured interviews were the main data source for this portion of the pilot study. In a semi-structured interview, a limited number of main interview questions and follow-up questions are prepared in advance (Rubin & Rubin, 2012). Semi-structured interviews allow for specific questions to be queried, while simultaneously providing the opportunity for conversation with the participants in relation to their feelings and experiences about standardized testing environments (Creswell, 2007).

To enable each participant to articulate their experiences with standardized testing environments in their own words, questions were phrased in an open-ended format (Hatch, 2002). Open-ended questions were designed to capture each participant's experiences with standardized testing environments. For example, the participants were asked, "What feelings or emotions did you have during the test?" and "How did you feel about what was going on in the room during the test?" The interview also included prompting questions to help the contributor understand the structured question or keep the participant on topic. For example, when more information was required, participants were asked "Why?" or "How so?" Examples of the open-ended and prompting questions used during the interviews can be found in Appendix B.

Data Analysis: Pilot Study Phase II (Qualitative)

Hatch (2002) states that qualitative data analysis is the systematic search for meaning. In qualitative research, data analysis occurs simultaneously with data collection. Data analysis begins with the first dialogue (i.e., the initial interview; Merriam, 2002), or through reflection on the interview prior to transcription. The immediate and constant analysis allowed for adjustments and reflection from the beginning to the culmination of the research process. For example, the behaviors of fellow test takers were investigated after participants mentioned being distracted by the actions of others in the testing room.

Frequent memoing was also conducted throughout the study. Memoing is the act of writing notes about what is being learned from the data (Charmaz, 2014). Memoing included written reflections, asking questions, and taking notes on the data collection and analysis process. An example of a question that arose through memoing is, “Why did the participants have better experiences at certain locations and not at others?” This study used both informal (i.e., in-the-field) analysis and systematic (i.e., out-of-the-field) analysis. Both data analytic strategies are detailed in the following sections.

Informal Analysis: Pilot Phase II (Qualitative)

The informal analysis began with the initial interviews. In-the-field analysis included the selection of interview questions, providing clarification to interview questions, using “off-the cuff” questions that are asked as part of the natural flow of conversation during the interview, and probing for more detail. Copious notes were taken during the interviews and after the audio tapes were reviewed. Researcher comments were included in the interview notes.

To not solely rely on audio recordings, researcher comments were documented to stimulate critical thinking about what is seen and heard during the interview (Bogdan &

Biklen, 1998). These notes included reflections on the inclusion and omission of data. This informal data analysis helped to shape the decision that saturation was reached and that the inclusion of additional participants was unnecessary. Saturation is defined as when the investigation reaches a point where no new information will be obtained from further interviews and data analysis (Teeter & Sandberg, 2016). After the fourth participant was interviewed, the participants' responses appeared to be restatements of the same ideas and no new themes developed.

A researcher's journal was maintained to informally analyze the data. This journal was used throughout the study, from pre-interviews to conclusions, for reflection. The process of frequently reflecting on the data is known as iteration. Iteration is the process of repeating rounds of exploration with the goal of gaining insight and meaning after each sequence (Srivastava & Hopwood, 2009). The data were scoured on multiple occasions to allow themes to develop that expressed the meaning behind the interviewees intentions. The journal was a record of the interviews, reflections, errors, ideas, or epiphanies that developed during the study.

Systematic Analysis: Pilot Phase II (Qualitative)

In the systematic data analysis portion of the study, the process of inductive analysis was used. Inductive data analysis takes specific pieces of information and organizes them together as a meaningful whole (Hatch, 2002). Using inductive data analysis (i.e., inductive reasoning), patterns in the data were identified that organized how recent high school graduates experience standardized assessments. After systematically studying, organizing, and categorizing the data, the categories were then coded. Coding is defined as giving labels to portions of the data that allow for the classification and synthesis of information, while still accounting for all components of the data (Charmaz, 2014). These codes were not predetermined, but rather a result of allowing patterns to develop from the data.

The data were coded in larger sections of sentences, which allowed for immersion without becoming preoccupied with individual words (Thorne, Reimer, Kirkham, & MacDonald-Emes, 1997). One phase of coding was involved in an effort to develop themes and identify patterns in the data, consistent with a basic interpretive design. Data analysis was completed when the the following questions were addressed: (1) "Can the analysis be explained and justified?," "Can a complete story be told?," and "Can the analysis be organized into coherent written findings?" (Hatch, 2002, p. 150).

Trustworthiness: Pilot Study Phase II (Qualitative)

Trustworthiness ensures that the study will have credibility with other researchers, policy makers, practitioners, but also with the research participants. The concept of trustworthiness is how the integrity of the study is maintained (Schram, 2006). As there appears to be a lack of research on how assessment environments are experienced by Millennial-aged individuals, trustworthiness is extremely important in this study. Several approaches were used to build trust in this study. Member checking was used to verify that the understanding of the participants' experiences was correct. Member checks involve asking the participants to comment on the interpretation of the data (Merriam, 2002). This was done during the interview process by asking questions such as, "So, what I am hearing is…?" and "Is my understanding of what you said correct?" Peer review was used as a measure to build trustworthiness such as regular consultation with advisors in the research community. Peer review was conducted through ongoing discussions and messages with advisors, instructors, and peer researchers communicated via a university-sponsored course online collaboration platform.

Additional meetings with research advisors on several occasions were conducted to discuss the data and interpretation. This ongoing feedback, and additional perceptions provided, increased the trustworthiness of this study. An audit trail was maintained, which describes in detail how the data were collected, how coding was derived, and how decisions were made throughout the study (Merriam, 2002). The audit trail included handwritten notes from each of the four interviews, explanations of the coding, evolution of the interview process, as well as notes on the overall themes that emerged

Reflexivity was also used to be critically self-aware of bias and perceptions. Since the researcher gathers the data, it is necessary to acknowledge personal biases and not allow them to influence the study or interpretations of the data. The researcher was aware of personal orientation related to standardized assessment environments. That is, quiet or silent environments are not preferred, and such conditions lead to personal distraction and a wandering mind. Every time the researcher has taken a standardized test, the environment was quiet without any stimuli. The investigator's personal bias is that the "universal model" for standardized testing may be flawed for some students. This personal bias was omitted from the research process by not sharing this information with the participants.

Ethics: Pilot Study Phase II (Qualitative)

Several procedures were used to promote ethics in this part of the pilot study. Confidentiality was maintained in every step of the research process. Pseudonyms were used when reporting findings. The pseudonyms were chosen by the participants and used to protect anonymity of the participants. Informed consent forms were signed by all participants, which detailed the purposes, procedure, data collection methods, risks, confidentiality/privacy, voluntary participation, and researcher's contact information. Participants were informed that they may stop participation at any time. Member checks were used to allow participants' input into the data analysis process. All data were kept in a secure location. At the conclusion of the study, all participants who requested were provided with a report of the initial findings.

CHAPTER IV (PILOT STUDY): RESULTS

Research findings presented in this chapter consist of cross-validation techniques to answer the research question "What are the psychometric properties (i.e., content and construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a university student population?" The first part of this chapter includes the Exploratory Factor Analysis (EFA) used to generate the theoretical factor model. A Confirmatory Factor Analysis (CFA) was used to verify the model. The following findings are reported: (1) Sample demographic information (prior to randomly splitting the data file), (2) Outliers and assumptions, (3) EFA demographics, (4) EFA item descriptive statistics and correlations, (5) EFA main analysis, (6) CFA demographics, (7) CFA item descriptive statistics and correlations, and (8) CFA main analysis.

Analysis Sample Demographic Information: Pilot Study Phase I (Quantitative)

Blanchard and Osborne (2010) recommend that investigators be aware of random responses from research participants. Study participants who are not interested in the outcome may provide random answers that are a threat to the validity of research. In order to protect against random responses, the data were screened and resulted in the elimination of invalid responses. A total of 666 undergraduate university students completed the TEOP. Items in the demographic portion of the measure included gender, race, cumulative Grade Point Average (GPA), ACT score, age, number of people in the household, number of children, parents' highest levels of education attained, and participants' preferred time of day for testing. Demographic and other characteristics of the final analysis sample are provided in Table 3.

Twenty-four participants were removed for being outside of the study's age demographic. Fifteen were eliminated for not answering the TEOP items, and five were dropped for entering an unrecognized GPA value on the scale. Two cases were eliminated for earning an undergraduate degree at the time of data collected. Finally, individual cases were dropped for unusual responses to the number of people living in their household, for not taking the ACT, and for describing their gender as an "Apache Attack Helicopter." After removal of cases, a total of 617 participants remained. These cases were randomly split into two samples containing 308 (i.e., the EFA sample) and 309 (i.e., the CFA sample) participants.

Outliers and Assumptions: Pilot Study Phase I (Quantitative)

Extreme values (i.e., outliers) are cases that fall outside the normal range on one or more variables. These outliers can cause a distortion (e.g., skewness) in the data set. The presence of an outlier can lead to an increased risk of committing a Type I or Type II Error. Furthermore, outliers make the generalizability of the results minimal at best (Lomax & Hahs-Vaughn, 2012). In order to conduct FA, certain multivariate assumptions must be met. These assumptions include: (1) Normality, (2) Factorability, and (3) Sample Size. Normality is observed when the distribution of each observed variable (i.e., the TEOP items) is normal (i.e., univariate normality). Secondly, factorability is the assumption that there are at least some correlations amongst the variables to identify factors. Lastly, the sample size should be large enough to yield reliable estimates of correlations among the variables.

Normality was checked prior to conducting the EFA by examining skewness for each item. Five of the eleven items were significantly skewed ($p < .001$ for all). Factorability was checked using the Kaiser-Myer-Olkin (KMO) Test of Sampling Adequacy and Bartlett's Test of Sphericity. The KMO value should be greater than .60 and Bartlett's Test of Sphericity should be significant (Neil, 2017). The value of KMO was acceptable (.719) and Bartlett's was significant ($p < .001$). Finally, the sample size should be large enough to yield reliable estimates of correlations among the variables. Nunnally (1994) suggests that 10 subjects per item is necessary to reduce sampling error. Thus, the sample size of 308 participants was considered acceptable.

Exploratory Factor Analysis (EFA) Descriptives: Pilot Study Phase I (Quantitative)

In the EFA sample ($N = 308$), 161 (52.3%) were male and 147 (47.7%) were female with a mean age of 20.52 ($SD = 1.85$). The EFA sample contained 249 (80.8%) White/Caucasian participants. The average ACT score was 22.85 ($SD = 3.64$), and the mean GPA was 2.96 ($SD = .85$). A Bachelor's degree was earned by 97 (31.5%) of the participants' mothers and 94 (30.5%) of their fathers'. One hundred and ninety-one (62.0%) of the respondents preferred taking a test between 11:01 am and 3:00 pm.

Table 3

Exploratory Factor Analysis Sample Pilot Study Phase I Demographics (*N = 308*)

Variables	*M*(*SD*) or *n*(%)
Age (*N* = 307)	20.52(1.85)
ACT Total Score (*N* = 246)	22.85(3.46)
Grade Point Average (GPA)	2.96(.85)
Gender	
Male	161(52.3)
Female	147(47.7)
Ethnicity	
White/Caucasian	249(80.8)
Black/African-American	18(5.8)
American Indian/Alaskan Native	2(.6)
Asian	10(3.2)
Hispanic/Latino	9(2.9)
Bi or Multiracial	11(3.6)
Other	9(2.9)
Mother's Education	
Not Sure	7(2.3)
High School Diploma	72(23.4)
Some College	69(22.4)
Bachelor's	97(31.5)
Master's	45(14.6)
Professional	12(3.9)
Doctorate	4(1.3)
Missing	2(.6)
Father's Education	
Not Sure	6(1.9)
High School Diploma	90(29.2)
Some College	62(20.1)
Bachelor's	94(30.5)
Master's	43(14.0)
Professional	9(2.9)
Doctorate	4(1.3)
Preferred Testing Time	
7:01 AM – 11:00 AM	40(13.0)
11:01 AM – 3:00 PM	191(62.0)
3:01 PM – 7:00 PM	61(19.8)
7:01 PM – 11:00 PM	11(3.6)
11:01 PM – 3:00 AM	3(1.0)
3:01 AM – 7:00 PM	1(.3)
Missing	1(.3)

Item Descriptives and Correlations

In the EFA sample, descriptive statistics for all items are found in Table 4. Item 1 (i.e., “I prefer silence.”) was reverse coded to allow for all items in the “Sound” factor to be negative. The item with the highest mean was Item 1 (i.e., “I prefer silence.”; M = 4.41, SD = .80), and the item with the lowest mean was Item 4 (i.e. “I prefer to have a television on in the background.”; M = 1.83, SD= 1.21). Inter-item correlations (Allen & Yen, 2002) were used to examine the relationships between items. Of the significant correlations, the relationship between Item 8 (i.e., “I prefer to have the option to stand up.”) and Item 9 (i.e., I prefer to have the option to walk around.”) was the highest (r = .643, $p < .001$). The lowest correlation was between Item 4 (i.e., “I prefer to have a television on in the background.”) and Item 10 (i.e., “I prefer to take it on a computer.”; r = .121, $p < .05$). Item 11 (i.e., “I prefer to use pencil and paper.”) was only significantly correlated with one other item (Item 10: “I prefer to take it on a computer.” r = -.123, $p < .05$). The correlations between all the items are presented in Table 5.

Table 4

Exploratory Factor Analysis Sample Pilot Study Phase I TEOP Item Descriptive Statistics (N = 308)

Items	*M*	*SD*	*Mdn*	IQR	Skew	Kurt
1. I prefer silence.	4.41	.80	5.00	1	-1.63	3.51
2. I prefer background/noise environmental sounds.	2.73	1.37	3.00	3	-.29	-1.57
3. I prefer to listen to music.	2.34	1.41	3.00	3	.306	-1.52
4. I prefer to have a television on in the background.	1.83	1.29	1.00	2	1.03	-.70
5. I prefer to have water/a beverage available.	3.93	.99	4.00	2	-.99	1.21
6. I prefer to have food available.	3.04	1.28	3.00	1	-.59	-.92
7. I prefer to chew gum/candy.	3.47	1.13	4.00	1	-.89	.37
8. I prefer to have the option to standup.	2.91	1.39	3.00	3	-.39	-1.37
9. I prefer to have the option to walk around.	2.74	1.37	3.00	3	-.24	-1.48
10. I prefer to take it on a computer.	3.36	1.01	3.00	1	-0.71	-0.71
11. I prefer to use paper and pencil.	3.57	.89	4.00	1	-0.49	1.03

Note. Min = 1 and Max = 5 for all items.

Table 5

Exploratory Factor Analysis Sample Pilot Study Phase I Inter-Item Spearman Correlation Matrix for the TEOP Items (N = 308)

Items	1	2	3	4	5	6	7	8	9	10	11
1. Silence	-	.292***	.367***	.391***	-.047	.153**	.050	.073	.104	.008	-.073
2. Background Noise		-	.348***	.240***	-.042	.118**	.010	.127*	.145*	.087	.028
3. Music			-	.543***	.071	.250***	.160**	.168**	.263***	.090	-.042
4. Television				-	.018	.241***	.144**	.227**	.302***	.121*	-.049
5. Water/Beverages					-	.182**	.155**	.126*	.153**	.142*	.018
6. Food						-	.326***	.385***	.395***	.207***	.010
7. Chew Gum/Candy							-	.200***	.168**	.145*	.006
8. Stand Up								-	.643***	.124*	.041
9. Walk Around									-	.137*	.043
10. Computer										-	-.123*
11. Paper and Pencil											-

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Exploratory Factor Analysis (EFA): Pilot Study Phase 1 (Quantitative)

In this study an EFA, specifically Principal Axis Factoring (PAF), was used to investigate the factorability of the 11 TEOP items. First, it was observed that multicollinearity was not a concern as there were no inter-item correlations greater than

.8 using a Spearman correlation matrix (Pett, Lackey, & Sullivan, 2003). Second, as stated previously, the KMO measure of sampling adequacy was .719, which was greater than the commonly recommended value of .6 (Pett et al., 2003). Third, Bartlett's Test of Sphericity was significant (χ^2 [55] = 670.271, $p < .001$), indicating correlations were sufficiently large to conduct FA.

PAF was selected as the extraction method with Direct Oblimin rotation. The EFA on the TEOP items revealed three factors with an eigenvalue > 1.00, which explained 53.603% of the variance. The scree plot showed that these data had two factors. Researchers suggest that if the number of factors to extract is unclear, several FAs with differing numbers of specified factors can be run (Gorsuch, 1983, Tabachnick & Fiddell, 2001). Two- and three-factor extractions were examined. Nunnally and Bernstein (1994) state that usefeullness and interpretability should ultimately frame the decsion on the number of factors to extract. Based on statistical criteria and theory, two factors were extracted.

The pattern matrix (see Table 6) showed that the items loaded on two factors based on loadings of .32 or higher. Tabachnick and Fiddell (2001) state that .32 is a good rule-of-thumb for the minimum loading of an item on a factor, as it accounts for approximately 10% of the overlapping variance with the other items in that factor. Items with low loadings (< .32) were excluded from further analysis. Items 10 and 11 (i.e., "When taking a test, I prefer to take it on the computer." and "When taking a test, I prefer to use paper and pencil.") did not load on either factor. Items 5 through 9 all loaded on Factor 1 and involved an act or behavior. This factor was called "Action." The highest loading for the “Action” factor was Item 8 (i.e., “When taking a test, I prefer to have the option to stand up.”) with a loading of .764. The lowest was Item 5 (i.e., “When taking a test, I prefer to have water/beverage available.”) with a loading of .325. The “Action” factor accounted for slightly more than 27% of the variance, and the item correlations within this factor (i.e., Items 5-9) were all significant ($p < .05$ for all).

Items 1 through 4 loaded on Factor 2 and all involved hearing and sound. This factor was named “Sound” and accounted for over 14% of the variance. Item 3 (i.e., “When taking a test, I prefer to listen to music.”) had the strongest loading at -.706. The lowest loading on the “Sound” factor was Item 2 “When taking a test, I prefer background noise/environmental sounds” with a loading of -.445. All items (i.e., Items 1-4) within this factor had significant correlations ($p < .001$ for all). Coefficient α was used as a measure of internal consistency reliability for the final nine-item TEOP (Cronbach, 1951). The total TEOP showed evidence of high internal consistency reliability ($\alpha = .730$). The five-item "Action" factor had a slightly higher reliability ($\alpha = .695$) relative to the four-item "Sound" factor ($\alpha = .685$). The correlation between the factors “Action” and “Sound” was significant ($r = .318$, $p < .001$).

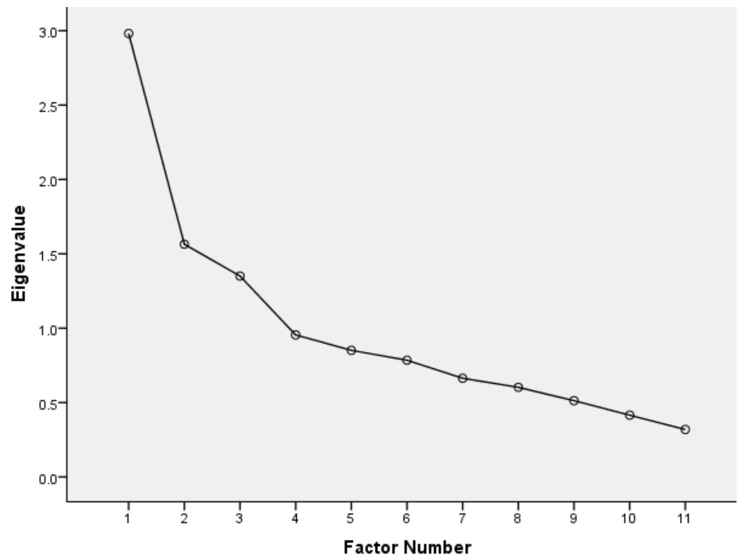

Figure 1. Scree Plot Exploratory Factor Analysis (EFA) Pilot Study Phase I.

Table 6

Exploratory Factor Analysis Sample Pilot Study Phase I Principal Axis Factoring with Direct Oblimin Rotation Item Loadings (N = 308)

Items		Factor	
		1	2
5.	I prefer to have water/a beverage available.	.325	
6.	I prefer to have food available.	.581	
7.	I prefer to chew gum/candy.	.347	
8.	I prefer to have the option to stand up.	.764	
9.	I prefer to have the option to walk around.	.730	
1.	I prefer silence.		-.564
2.	I prefer background noise/environmental sounds.		-.445
3.	I prefer to listen to music.		-.706
4.	I prefer to have a television on in the background.		-.634
10.	I prefer to take it on the computer.	-	-
11.	I prefer to use paper and pencil.	-	-

Confirmatory Factor Analysis (CFA) Descriptives: Pilot Study Phase 1 (Quantitative)

In the CFA sample (*N* = 309), 108 (35.0%) were male and 201 (65.0%) were female, with a mean age of 20.92 (*SD* = 2.58). The CFA sample was comprised of 261 (84.5%) White/Caucasian participants. The mean ACT score was 22.99 (*SD* = 4.02) and the mean GPA was 3.08 (*SD* = .97). A Bachelor's degree was earned by 79 (25.6%) of the participants' mothers and 85 (27.8%) of their fathers. One hundred ninety-three (62.5%) of the respondents preferred taking a test between 11:01 am and 3:00 pm.

Table 7

Confirmatory Factor Analysis Sample Pilot Study Phase I Demographics (*N = 309*)

Variables	*M*(*SD*) or *n*(%)
Age (*N* = 305)	20.92(2.58)
ACT Total Score (*N* = 256)	22.99(4.02)
Grade Point Average (GPA)	3.08(.97)
Gender	
Male	108(35.0)
Female	201(65.0)
Ethnicity	
White/Caucasian	261(84.5)
Black/African-American	21(6.8)
American Indian/Alaskan Native	1(.3)
Asian	1(.3)
Hispanic/Latino	6(1.9)
Bi or Multiracial	8(2.6)
Other	8(2.6)
Missing	3(.9)
Mother's Education	
Not Sure	9(2.9)
No High School Diploma	5(1.6)
High School Diploma	86(27.8)
Some College	81(26.2)
Bachelor's	79(25.6)
Master's	41(13.3)
Professional	3(1.0)

Doctorate	4(1.3)
Missing	1(.3)
Father's Education	
Not Sure	11(3.6)
No High School Diploma	8(2.6)
High School Diploma	108(35.0)
Some College	57(18.4)
Bachelor's	85(27.5)
Master's	24(7.8)
Professional	7(2.3)
Doctorate	8(2.6)
Missing	1(1.3)
Preferred Testing Time	
7:01 AM – 11:00 AM	48(15.5)
11:01 AM – 3:00 PM	193(62.5)
3:01 PM – 7:00 PM	49(15.9)
7:01 PM – 11:00 PM	7(2.3)
11:01 PM – 3:00 AM	10(3.2)
3:01 AM – 7:00 PM	2(.6)

Table 8

The Test Environment for Optimal Performance (TEOP) Items (N = 9)

#	Item
1.	When taking a test, I prefer silence.
2.	When taking a test, I prefer background noise/environmental sounds (e.g., sounds of the ocean, rain, birds chirping).
3.	When taking a test, I prefer to listen to music.
4.	When taking a test, I prefer to have a television on in the background.
5.	When taking a test, I prefer to have water/beverage available.
6.	When taking a test, I prefer to have food available.
7.	When taking a test, I prefer to chew gum or candy.
8.	When taking a test, I prefer to have the option stand up.
9.	When taking a test, I prefer to have the option to walk around.

Item Descriptives and Correlations

In the CFA sample, descriptives for all items are found in Table 9. The item with the highest mean was Item 1 (i.e. "I prefer silence."; $M = 4.11$, $SD = .97$), and the item with the lowest mean was Item 4 (i.e., "I prefer to have a television on in the background."; *M* = 1.67, *SD*= 1.04). Of the significant correlations, the relationship between Item 8 (i.e., "I prefer to have the option to stand up.") and Item 9 (i.e., "I prefer to have the option to walk around.") was the highest ($r = .747$, $p < .001$). The lowest correlation was between Item 4 (i.e., "I prefer to have a television on in the background.") and Item 7 (i.e., "I prefer to chew gum/candy."; $r = .114$, $p < .05$). The correlations between all the items are presented in Table 10.

Table 9

Confirmatory Factor Analysis Sample Pilot Study Phase I TEOP Item Descriptive Statistics (N = 309)

Items	*M*	*SD*	*Mdn*	IQR	Skew	Kurt
1. I prefer silence.	4.11	.97	4.00	2	-.90	.21
2. I prefer background/noise environmental sounds.	2.32	1.19	2.00	2	.37	-.96
3. I prefer to listen to music.	2.27	1.32	2.00	2	.58	-.94
4. I prefer to have a television on in the background.	1.67	1.04	1.00	1	1.37	.75
5. I prefer to have water/a beverage available.	4.00	1.05	4.00	2	.94	.51
6. I prefer to have food available.	2.91	1.32	3.00	2	-.01	-1.09
7. I prefer to chew gum/candy.	3.50	1.21	4.00	1	-.64	-.38
8. I prefer to have the option to standup.	2.61	1.30	3.00	3	.27	-1.04
9. I prefer to have the option to walk around.	2.59	1.35	2.00	3	.25	-1.24

Note. Min = 1 and Max = 5 for all items.

Table 10

Confirmatory Factor Analysis Sample Pilot Study Phase I (Quantitative) Inter-Item Spearman Correlation Matrix TEOP(N = 309)

Item	1	2	3	4	5	6	7	8	9
1. Silence	-	.505***	.524***	.445***	.070	.205***	.166**	.205***	.127*
2. Background Noise		-	.523***	.429***	.073	.237***	.153**	.179**	.146*
3. Music			-	.485***	.131*	.286***	.186**	.152**	.185**
4. Television				-	.036	.225***	.114*	.108	.101
5. Water/Beverages					-	.363***	.358***	.169**	.216***
6. Food						-	.425***	.313***	.341***
7. Chew Gum/Candy							-	.292***	.318***
8. Stand Up								-	.747***
9. Walk Around									-

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

The CFA was conducted using LISREL 9.10 (Joreskog & Sorbom, 2013). The results of the CFA were used to evaluate fit of the preliminary factor structure and to assess the underlying factor structure of the instrument. The intention of the factor analytic approaches in both the EFA and the CFA is to generate and validate a factor structure using the nine testing environmental preferences found in the TEOP. Inter-item correlations, the factors, and the factor loadings were examined to evidence the underlying structure of environmental preferences represented by the items in the measure.

CFA Main Analysis Pilot Study Phase I (Quantitative)

In the CFA results are presented in Table 11. For the inter-item correlations, Items 1 through 4 (i.e., the “Sound” factor) were all significantly and positively correlated with each other ($p < .001$ for all). The correlations between items within the “Action” factor (i.e., Items 5-9) were all significantly and positively correlated with each other as well ($p < .01$ for all). The initial CFA showed the observed variables water, food, gum, stand,

and walk all loaded significantly ($p < .001$ for all) on the factor "Action." Item 9 (i.e., "When taking a test, I prefer to have the option to walk around.") had the strongest loading (β = .87; 75.9% of the variance explained) and Item 5 (i.e., "When taking a test, I prefer to have water/a beverage available.") had the weakest loading (β = .28; 7.5% of the variance explained).

The observed variables silence, noise, music, and TV all loaded significantly ($p < .001$ for all) on the factor "Sound," with Item 3 (i.e., "When taking a test, I prefer to listen to music.") having the strongest loading (β = -.75; 56.5% of the variance explained) and Item 4 (i.e., "When taking a test, I prefer to have a television on in the background.") with the weakest loading (β = -.63; 39.4% of the variance explained). In the initial model there were nine factor loadings, nine measurement errors, and one factor correlation. Because the distinct values (i.e., unique values) in the matrix *S* (45) are greater than the total number of free parameters (19), this initial model is considered over-identified (i.e., there is more than one way of estimating parameters; (Schumacker & Lomax, 2016).

The χ^2 Goodness-of-Fit statistic was significant (χ^2 [26] = 112.33, $p < .001$). From the other fit indices, the Root-Mean-Square-Error of Approximation (RMSEA) was .104, the Standardized-Root-Mean Residual (SRMR) was .086, the Goodness- of- Fit Index (GFI) was .916, and the Adjusted Goodness-of-Fit (AGFI) Index was .854. Overall, the model did not fit the data. The modification indices were consulted, and a second model was run that allowed the error covariance of Item 8 (i.e., "I prefer to have the option to stand up.") and Item 9 (i.e., "I prefer to have the option to walk around.") to correlate. Additionally, these two items were located within the same "Action" factor, and conceptually, the behaviors of standing and walking are related actions (Jensen, 1998; Schumaker & Lomax, 2016).

The final CFA model showed the observed variables water, food, gum, stand, and walk all loaded significantly ($p < .001$) on the factor "Action," with Item 6 (i.e., "When taking a test, I prefer to have food available.") having the strongest loading (β =.71; 51.0% of the variance explained), and Item 8 (i.e., "When taking a test, I prefer to have the option to stand up.") with the weakest loading (β = .44; 19.4% of the variance explained). The factor "Action" had a direct positive effect on preference for water, food, gum/candy, standing up, and walking around when testing. This predicts that as a participant's "Action" preference increases their inclination toward having water, food, gum/candy, as well as being able to stand up or walk around during testing will also increase. The strong loading of preference to have food available within the "Action" factor suggests that eating is seen as a deed and participants with greater scores on the "Action" items where inclined to prefer accessibility to food. The weaker loading on the preference to stand up indicates that a lower score correlation between this item and the rest of the "Action" items. This may be the result of the error covariance between Item 8 and Item 9 being allowed to correlate.

The observed variables silence, noise, music, and TV all loaded significantly ($p < .001$) on the factor "Sound", with Item 3 (i.e., "When taking a test, I prefer to listen to music.") having the strongest loading (β = -.76; 57.6% of the variance explained), and Item 4 (i.e., "When taking a test, I prefer to have a television on in the background.") having the weakest loading (β = -.63; 39.4% of the variance explained). The factor

"Sound" had a direct negative effect on preference for silence, background noise, music and television when testing. This predicts that as a participant's "Sound" preference decrease their penchant for having silence, background noise, music, and television on during testing will increase. The strong loading for a preference of listening to music when testing suggests that those with a high sound preference had higher scores on this item. This could be due to music being strongly identified with sounds and makes conceptual sense. The weaker loading on the preference for having a television on could be due to visual element provided by a television not present in the other items. The negative loading scores on the "Sound" items explain that as the participants preference for more " Sound" increases their scores on these items decreases. In the final model there were nine factor loadings, nine measurement errors, one factor correlation and one error covariance. Because the distinct values (i.e., unique values) in the matrix *S* (45) are greater than the total number of free parameters (20), this final model is considered over-identified.

The χ^2 fit statistic was not significant ($\chi^2[25] = 23.25, p > .05$). In addition, the RMSEA was .000, the SRMR was .033, the GFI was .983, and the AGFI was .970. Overall, all model fit indices suggested a good fit and there were no additional modifications (Schumaker & Lomax, 2016). The two-factor structure of the TEOP used in the CFA produced evidence of high internal consistency reliability ($\alpha = .767$). The four-item "Sound" factor had a slightly higher reliability ($\alpha = .765$) compared to the five-item "Action" factor ($\alpha = .743$). There were negative correlations between the two subscales (i.e., "Sound" and "Action") in both the initial model ($r = -.287, p = .064$) and the final model (*r* = -.411, *p* =.067). This slight negative correlation indicates an inverse relationship between "Action" and "Sound" scores, though it is not statistically significant.

Table 11

Pilot Study Phase I (Quantitative) Standardized Factor Loadings for the TEOP(N = 309)

Item	Initial Model		Model 1	
	F	β	F	β
I prefer to have water/a beverage available.	Action	.28	Action	.50
I prefer to have food available.	Action	.43	Action	.71
I prefer to chew gum/candy.	Action	.40	Action	.62
I prefer to have the option to stand up.	Action	.84	Action	.44
I prefer to have the option to walk around.	Action	.87	Action	.48
I prefer silence.	Sound	-.71	Sound	-.70
I prefer background noise/environmental sounds.	Sound	-.70	Sound	-.70
I prefer to listen to music.	Sound	-.75	Sound	-.76
I prefer to have a television on in the background.	Sound	-.63	Sound	-.63

Note. F = Factor for each item (i.e., Action and Sound); β = the standardized factor loadings; Initial Model = the original 2-factor; Model 1 = Error covariance was added between Item 8 and Item 9; CFA= Confirmatory Factor Analysis.

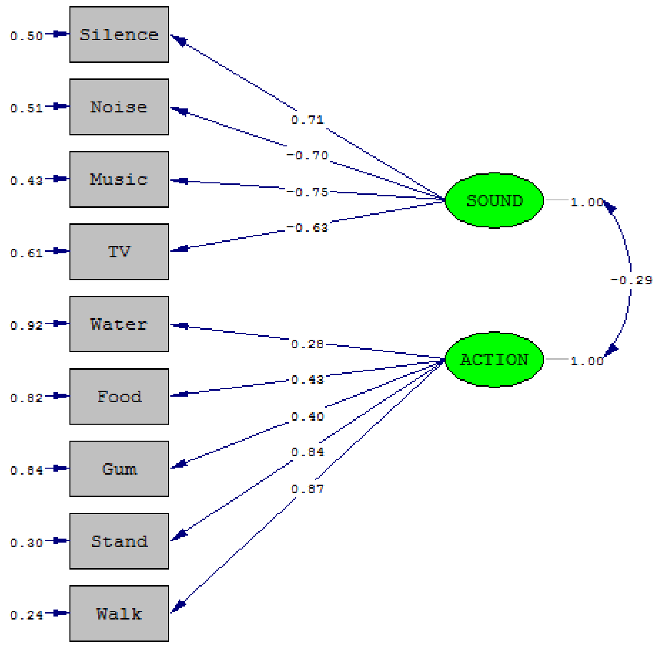

Figure 2. Initial model for the Confirmatory Factor Analysis (CFA) Pilot Study Phase I.

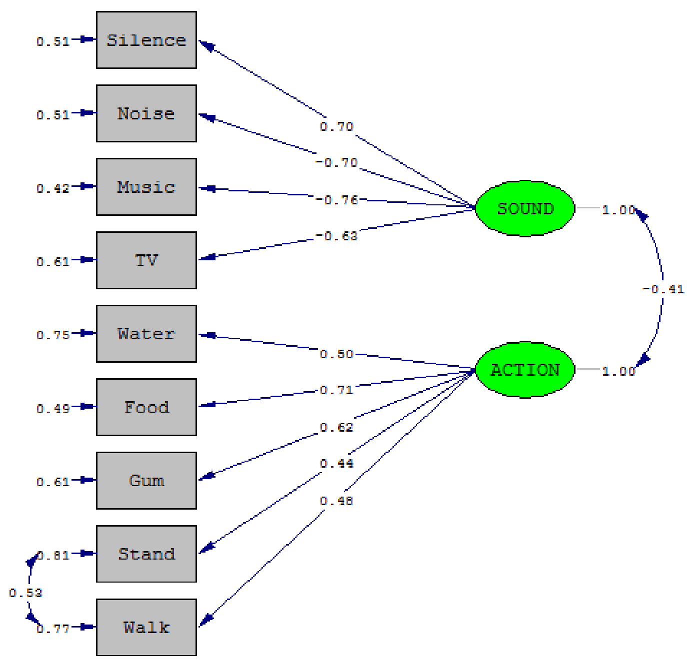

Figure 3. Final model for the Confirmatory Factor Analysis (CFA) Pilot Study Phase I.

Findings: Pilot Study Phase II (Qualitative)

The four participants were all 2016 or 2017 high school graduates who are either currently attending college or starting in the fall of 2017. The participants were 18 or 19 years old at the time of the interview. Two of the participants were male and two were female and from the same Midwestern state. The students all had higher scores on the ACT (i.e., ranging from 24 to 29) compared to the average score ($M = 20.8$) for the general student population in the US (ACT, 2017).

Table 12

Demographic Profile of Participants Pilot Phase II (Qualitative)

Name	Age	Sex	ACT Date	ACT Score	Hobby	Major
Chrissy	18	F	02/17/17	28	Softball	Actuarial Science
John	19	M	06/15/16	27	Lacrosse	Criminal Justice
Moses	18	M	05/16/17	29	Powerlifting	Pre-Medicine
Anna	18	F	05/16/17	24	Volleyball	Communications

Themes: Pilot Study Phase II (Qualitative)

Five themes captured the perception Millennial standardized test-takers have of testing environment in this study. These themes included the following: (1) the pretest experience at the testing location, (2) the testing room, (3) personal actions and the actions of others during the test, (4) sounds during the test, and (5) ideal testing environments. The following sections will provide a description of the participants' observations of their experiences in relation to the five themes. The descriptions of these themes assist in addressing the research question, "What are recent high school graduates' perceptions of their experiences with current standardized testing practices and environments?"

Theme 1: Pretest Experience at Testing Location

Because of the high-stakes nature of the ACT, the participants expressed an anxiety going into the test. Participants mentioned that having to wait to start the test was a point of frustration. Chrissy noted, "I arrived at the time I was told and had to wait for a half hour in the cafeteria prior to the test." This wait time was a point of frustration for Chrissy and it increased her level of anxiety. Anna had to wait an hour from when her test was scheduled until the start of the test. Anna said, "I felt prepared to answer the material but the atmosphere made me nervous and the longer I had to sit and wait the worse it got. I get test anxiety when taking standardized tests and math tests, so the wait definitely didn't help." John also had to wait approximately 30 minutes before the start of the ACT, and while it did not make him nervous it was a point of frustration: "Waiting was really annoying; I mean they told me when to get there and then they weren't ready. I just wanted to get in and get the test over with." In contrast with the others, Moses did not have to wait to start the test. He said the process took approximately ten minutes.

The moderators' behaviors prior to the test were important to the participants in how they perceived the environment. Chrissy felt supervisors were rude and cranky. She consequently had a very negative experience at that testing location. Anna took the ACT at two different locations and noted that the moderators were extremely rude at one location and kind at the other. She stated, "The administrator was so sweet. This really helped me relax." Moses knew the test supervisors as they were friends of his father, and he said they really were not noticeable.

Theme 2: Testing Room

The participants took the test at various locations in several communities in the Great Lakes region of the US. The test locations were high schools in the community as well as at an urban state university that serves primarily commuters. The ACT was administered at high schools in classrooms in which some had windows and some did not. All classrooms had posters and other adornments typically seen in high school classrooms. Depending on the test site, participants were in classrooms with approximately 15 to 20 other test takers at the same time. A large lecture hall was the testing room for the ACT that was administered at the university, with between 40 to 50 test takers in one group. The lecture hall did not have any decorations on the walls.

At the testing location, the students were assigned a seat and not allowed to move. Anna was fortunate that she was assigned near the windows. She added, "It was nice to be able to look out and look for birds in the trees when I needed a mental break. I wish the windows would have been open though because hearing the birds and feeling a breeze would be nice." Moses, who took the test at both the lecture hall and in a high school classroom, preferred the classroom setting. He noted, "The classroom setting had windows and I could look out and see birds and sunshine." Additionally, Moses mentioned that the posters "gave me something to look at." John's testing location did not have windows, but he stated that it would have helped him if it did: "Sometime my mind wonders when testing and I need a distraction." Chrissy indicated that her room had windows, but she did not sit near them. “I really focus when testing and I am very concerned about time limits, so I do not want to look up from the test.” Overall, all four participants took the test multiple times and at multiple sites, and each had a favored location. Coincidentally, their preferred environment was also the location that resulted in their highest score on the ACT.

Theme 3: Actions During Test

The permitted and prohibited behaviors of the participants as well as the actions of the other test takers in the location were a point of emphasis among those interviewed. They mentioned how others in the room triggered distractions and fear during the ACT. Additionally, the participants noted that certain actions/behaviors that they were or were not allowed to do impacted their performance. One collective experience was that all four interviewees were very aware of the others in the testing room.

When other test takers turned pages and finished sections prior to the participants it was noticeable. Anna was particularly impacted by what she referred to as a child prodigy who took the ACT with her: “He wasn’t even in high school yet and finished super early every time.” She was referring to him finishing the sections of the ACT. Anna noted how distracting that was by stating, “I’m like I don’t have enough time and he’s done already.” She said this made her feel rushed which in turn, made her also feel inadequate as others in the room turned pages or finished before her. This feeling of insufficiency was shared by all participants four in the study. Additionally, Anna was distracted by other test takers tapping pencils and/or feet, and especially those who read parts of the test out loud. She said, “I’m like shut up, I am trying to take my test.”

Moses preferred the test room with less people to the large lecture hall: “I would prefer to test by myself. I get weird about other people looking at me doing stuff. I had to sit in the front of the room and was concerned the whole test that others were looking at me.” Moses commented that he was aware of the others in the testing room and what they were doing. He indicated, “Other people turning pages is very noticeable to me. I wonder if I am going to have enough time, or am I smart enough.” It should be noted that Moses was concerned about being “smart enough” when taking the test. However, Moses had the highest ACT score of the four participants, and had already been accepted into Medical School following an accelerated undergraduate program. Although Moses experienced a history of academic success, he felt apprehensive that he was not competent with regards to taking the ACT.

Chrissy was also aware of the others finishing before her. Chrissy’s primary concern was finishing in the allotted time, and noted that when others finished before her, it made her feel rushed. Chrissy was also distracted by the supervisors walking around the room.

The study participants mentioned that they had preferences for certain actions that were forbidden according to the testing rules. These include having access to food, water, and gum. Additionally, some of those interviewed would have liked the ability to move around during the test to clear their minds. John mentioned how he does not eat breakfast and he tended to slow down at the end of the test. He said, “Having a snack would have provided a nice pick me up.” This feeling was mirrored by Moses who got hungry during the test: “Toward the end of the test I was real hungry. I think I would have done better if

I could have brought a sandwich or a snack." Anna and Chrissy both felt that having food would have only been a distraction. They both noted that food would slow them down, and would exacerbate their concern about the time limit.

Moses stated that having water or something available to drink during the test would have helped. He noted, "I get bored when taking tests and having water available would help break up the monotony and allow me to refocus." John said he frequently got dry mouth during the test, and the thirst caused his mind to drift. He said, "I haven't had a drink in a while. If I could have had water around it would have eliminated that distraction." John also thought having access to water during the test would have helped him with what he called "the overwhelming questions." He mentioned, "I could take a quick drink and refocus." Chrissy felt that not having water did not impact her, but she mentioned that she typically has water with her when she studies.

Anna felt particularly disadvantaged by not being allowed to take water with her into the testing room. Anna takes water everywhere and noted, "I like having water with me at all times. I study with it and carry it everywhere. It was frustrating not having it there for the ACT, if you are sitting there a drink would refresh you, particularly when reading the same passage over and over, plus classrooms get hot and having something to drink would be a benefit." Anna was uncomfortable without having water during the test because it was viewed as a source of security. Not having water was also coupled with not being allowed to bring gum into the testing center. Anna said, "I love chewing gum. Not having it during the test definitely distracted me. The constant motion of gum chewing helps my concentration. Plus, it tides me over when I get the munchies." Anna said emphatically, "I definitely would have done better if I could have chewed gum."

Chrissy was the most time conscious of the four students in the study and she mentioned that having the option to walk around would have simply cost her time to answer the questions. Moses felt that while normally he likes to move around, he did not really feel the need to do so during the ACT. Anna believed that it would have helped her when she felt frustrated to be able to get up and take a quick walk before returning to the test. John felt suffocated being in the same room for the duration of the test. He noted, "Sitting in the same spot for a long time is hard and it caused my mind to wander. I would have liked to get up and move around. Being able to stretch would have helped me to refocus on the test."

Theme 4: Sound During Testing

Chrissy noted that she is easily distracted during testing and prefers silence, as that is how she typically studies. Any background noise or music played in the room or having a headset would distract her and be a detriment to her performance. Anna said her sound preference would be to have environmental sounds such as birds, the ocean, or instrumental music. She also noted that she would sing along to songs with lyrics if music is played. Anna mentioned that one of her high school teachers played environmental sounds during unit tests and felt it was very beneficial. She noted, "It helped calm me down so much, the environmental sounds really helped my test anxiety."

John said that listening to music, particularly his own playlist, helps his test performance. He stated, "I study with music and when doing homework, so this would be helpful during the ACT." John was clear that he would not want to listen to others' music. The ideal situation for John would be to have music of his choosing playing in the background during testing, so it more closely mimics how he studies. He indicated, "Having headphones for music would be great."

Moses said while he studies with headphones on he does not think having music played during the ACT would impact his score. Moses said that while he does not like "fake" environmental sounds since they are unnatural, "It's not realistic to hear waves crashing in a classroom, but I think being able to hear real birds chirping would help me though." Moses believes taking the ACT outside with authentic natural sounds would improve his performance.

Theme 5: Ideal Testing Environments

Each of the four participants in the study was asked to design an ideal test environment and describe what it might look like. Their ideal environments were unique to each person. The unique individual preferences diverge from the uniform testing environment used for the ACT and other major standardized tests. A commonality amongst the four participants was a preference for having less people in the testing room. Three of the four indicated preferring to take the test alone. Chrissy said that she would like the room to be soundproof and her to have water available. John identified that his ideal test environment would be to take the test alone, and would prefer to listen to his music with the room silent to all outside noise. He also noted that he would like both food and water available, with the option to move around as he sees fit, and working at his own pace. Similar to the other participants, Moses said his ideal environment would be to take the test alone. He indicated that he would prefer to be outside near a pond, with grass at his feet, while taking the test. Moses wants to hear the genuine sounds of nature. Finally, Anna’s ideal environment would be to test with a small group of 5 to 10 other test takers in a round room with open windows. She mentioned that the fresh air helps her concentrate, and the open windows would allow her to hear the birds chirping. Anna strongly prefers having both water and gum available to her during tests.

CHAPTER V (PILOT STUDY): DISCUSSION

Discussion (Pilot Phase I)

The objective of this study was to examine current college students' perceptions of preferred standardized testing conditions. The primary research question that guided the study was, "What are the psychometric properties (i.e., content and construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a university student population?"

Exploratory Factor Analysis (EFA)

In the first part of the pilot study, the EFA sample showed that the TEOP items loaded on two factors "Action" and "Sound." Two items, 10 and 11 (i.e., "When taking a test, I prefer to take it on the computer." and "When taking a test, I prefer to use paper and pencil.") did not load on either factor. The five items that involved an act or behavior loaded on the factor called "Action." The label "Action" is appropriate because drinking, eating, chewing, standing up, and walking around all involve the test taker actively engaging in a behavior. The highest loading on the "Action" factor was Item 8 (i.e., "I prefer to have the option to stand up.") and the lowest was Item 5 (i.e., "I prefer to have water/a beverage available."). For Item 8, the process of standing up from a seated position involves movement (i.e., an action). And for Item 5, the wording suggests that having water/a beverage available during testing includes the action of drinking the liquid. However, drinking is not often seen as a primary action, but rather an innocuous behavior. Item 5 had the highest score amongst the action items in the EFA sample, but there was minimal discrimination for this item between the "Action" and "Sound" factors. This resulted in the lowest loading among the Action items. Item 9 (i.e., "I prefer to have the option to walk around.") had the lowest score among the "Action" items.

The four items that involve hearing and audible noise loaded on a factor titled "Sound." The factor name "Sound" is appropriate because the four items all relate to the test takers' preferences regarding noise. Item 3 (i.e., "I prefer to listen to music.") had the strongest loading on "Sound." This makes conceptual sense as listening to music is, by definition, a sound preference. The lowest loading on the "Sound" factor was Item 2 (i.e., "I prefer background environmental noises."). The lower loading is most likely due to the word "background" in the item, which implies that the sound is more censored and unobtrusive. Item 1 (i.e., "I prefer silence.") had the highest score, and Item 4 (i.e., "I prefer to have a television on in the background.") had the lowest score. These findings indicate that most participants prefer silence, and few if any, like to have distracting and/or slightly audible background television noise when testing. Finally, for the two-factor TEOP structure, the five-item "Action" factor had slightly higher internal consistency reliability relative to the four-item "Sound" factor. Although only a difference of one item, if there are more items in a scale/measure, there is a commensurate increase in internal consistency reliability (Cortina, 1993).

Confirmatory Factor Analysis (CFA)

A CFA was conducted to test if the data fit the hypothesized factor structure. After consulting the modification indices, the model included the observed variables of water, food, gum, stand, and walk all loading significantly on the "Action" factor. Item 6 (i.e., "I prefer to have food available.") had the strongest loading and the "Action" factor explained the largest proportion of variance in this item. This can be attributed to the deliberate steps necessary to bring food to the testing location. Additionally, Item 8 (i.e., "I prefer to have the option to stand up.") had the weakest loading and the "Action" factor explained the least amount of variance in this item. Item 8's loading decreased considerably from the initial model to the final model. This reduction can be attributed to the added error covariance between Items 8 and 9.

The observed variables of silence, noise, music, and TV all loaded significantly on the factor "Sound." Item 3 (i.e., "I prefer to listen to music.") had the strongest loading and the "Sound" factor explained the largest proportion of variance in this item. The strong loading can be explained by the direct association between music and sound. Item 4 (i.e., "I prefer to have television on in the background.") had the weakest loading and the "Sound" factor explained the smallest proportion of variance in this item. The low loading could be attributed to television being perceived as a visual medium more so than sound.

The largest score from the TEOP "Action" items in the CFA sample was Item 5 (i.e., "I prefer to have water/a beverage available.") and the lowest score was on Item 9 (i.e., "I prefer to have the option to walk around."). As indicated in the EFA section above, the high preference for being able to drink indicates that test takers would like the option of having a beverage during testing to perhaps help with concentration, calm anxiety, provide a break to think, and quench thirst. The lower preference for walking around may be due to the timed nature of these exams. Students may be less inclined to consider walking around during a test, as this would require moving away from their desk/computer station where the test is located. Students may be anxious or concerned that walking around will prevent them from completing the test on time. Finally, the highest mean score for “Sound” was Item 1 (i.e., “I prefer silence.”) and the lowest was Item 4 (i.e., “I prefer to have a television on in the background.”). As with the EFA sample, most participants prefer silence, and few favor noisy appliances/devices.

The inter-item correlations within the “Action” factor (i.e., Items 5 to 9) were all significant and positive in direction. The same was noted for Items 1 through 4 in the “Sound” factor. Additionally, stronger relationships were noted among items within the same factor compared to items located in the other factor. The strongest correlation among the “Action” items was between Item 8 (i.e., “I prefer to have the option to stand up.”) and Item 9 (i.e., “I prefer to have the option to walk around.”). Similarly for the “Sound” factor, Item 1 (i.e., “I prefer silence.”) and Item 3 (i.e., “I prefer to listen to music.”) had the strongest correlation. The two-factor TEOP structure was confirmed, with the four-item “Sound” factor having a slightly higher internal consistency reliability compared to the five-item “Action” factor. The CFA internal consistency reliabilities for the two factors were opposite in order of magnitude compared to the EFA reliabilities. For the CFA, the factor with fewer items had higher internal consistency reliability. Overall, the internal consistency reliability increased for both factors in the CFA.

Summary of Exploratory and Confirmatory Factor Analyses

Further examination of the item correlation matrices revealed a near “textbook” example of the within and between patterns of association that should be rendered after following best practices in construct-based measure construction. For “Action,” all the items contained within that factor had significant, positive, and high correlations with each other. These items were also not as strongly correlated with the items in the "Sound" factor. The correlations between items located in different factors (e.g., a correlation coefficient between an item in the "Action" factor and an item in the "Sound" factor) were inconsistent compared to the correlations amongst the items contained within each factor (e.g., a correlation coefficient between two items within the "Action" factor or between two items within the "Sound" factor). That is, there was less variation in the within-factor item correlations compared to the variation in the between-factor item correlations. The correlation patterns followed the expected template of between and within group relationships.

Discussion (Pilot Phase II)

In the second part of the pilot study, the purpose was to examine how recent high school graduates experienced standardized testing conditions, and to understand which environmental testing conditions may allow for optimal testing performance. The primary research question guiding this part of the study was: "What are recent high school graduates' perceptions of their experiences with current standardized testing practices and environments?" This qualitative portion of the pilot study was completed to provide support for the initial EFA and CFA findings. Few measurement-related studies involving the construct of test environment preferences exist. Thus, qualitative interviews were used to offer additional support for the factor structure. From interviews with four recent high school graduates, each participant had individual and unique experiences with standardized testing environments. These young adults also were specific in how they would design a standardized assessment environment specific to their own personal preferences.

The participants all took the ACT and their preferences were framed by their experiences related to taking this test. Additionally, all four students took the test multiple times (i.e., between two and nine) and in different locations. Each individual had a testing location that they preferred over the others for various reasons. Literature has noted that environmental conditions such as lighting and outside distraction have been shown to impact test outcomes (Zhu & Han, 2011). Additionally, personalized learning has also evidenced a positive impact on student performance (Kayluga & Sweller, 2005). Unsurprisingly, each student's favored site also produced his/her highest score on the ACT. As the preferred location produced better results, this implies that other preferences such as testing environment could also deliver improved outcomes.

Preferences

Participants had specific preferences for both "Actions" and "Sounds" during testing and within the testing environment. These individual preferences are in contrast to the "one-size-fits-all" model adopted by the ACT and other major standardized testing companies in which accommodating students' preferences are (generally) not permitted. Over their years in school, students have developed study habits, anxiety-reducing defense mechanisms, and other tools to help them achieve their optimal performance. For many students, the defined rules in administering high-stakes standardized tests do not permit them to use the abovementioned aides. The ideal test environment described by all four participants was very different from the crammed classroom or lecture halls used now. This aligns with the differentiated instruction techniques that were prominent in their formative education (Nunley, 2006; Subban, 2006). Millennials as a group prefer informal educational structure (Sweeney, 2006) This study's participants illustrate that as three participants preferred to test alone, one desired to test outside, and all four favored having water available.

Others Test Takers and Super Scoring

Students noted that they were acutely aware of others in the room and their behaviors during testing. The desire to reduce the number of people tested was common to all four participants, with three preferring to be alone. A particular concern was other students in the testing room turning pages and finishing a section early or before everyone else. This led to feelings of inadequacy and anxiety over the remaining test time. Zhu and Han's (2011) research informs that peripheral factors can have an affect on academic outcomes. Assessment companies such as the ACT and The College Board, in attempt to standardize conditions for all test takers, may inadvertently allow other possible personalized distractors to testing conditions, such as sounds and movements from the other individuals in the room (e.g., coughing, pencil tapping, others finishing early) that might negatively impact a participants performance.

One explanation for some students finishing these exams early (or too quickly) is the practice of "super scoring" the ACT, which is frequently used at college/universities. Super scoring involves taking the best individual score from each ACT section, regardless of the test date, for a new composite score (Seigel, 2016). As a result, some students taking the ACT are only concerned with one section on the exam. Those students allocate all their efforts to that one section and randomly select answers for the other sections. Super scoring is not endorsed by the ACT, but rather is a policy choice of individual universities (Wignall, 2017). Based on the responses from the participants in this study, they indicated being disturbed during testing by the behaviors of other students, and some of the noted behaviors were likened to the super scoring method.

Implications (Pilot Phases I and II)

Results highlight that Millennials have individual preferences for testing environment conditions. Examining the psychometric properties of the TEOP may be useful in understanding students' preferred testing environment to perform their best on an exam. A measure with valid and reliable scores can be beneficial to test takers who may be unaware of how testing preferences can impact their exam preparation and performance. For example, after completing the TEOP, students receive a score for both Action and Sound preferences. These scores are indicative of higher or lower preferences for Action, Sound, or both during testing. High scores on both Action and Sound suggest that students' preferences do not align with current high-stakes testing conditions, thus, students may need to adjust study strategies in order to be better prepared in the ACT testing environment. Additionally, students with high scores on either Action or Sound, but not both, can perhaps address the factor with the higher score in an attempt to be less dependent on those particular conditions during testing. Finally, students with low scores on both Action and Sound have preferences that are closely related to current high stakes testing environments. Overall, knowledge of students' testing preferences may encourage educators to emphasize preparation strategies that do not deviate from the current high-stakes standardized testing conditions. This emphasis of synchronizing of study habits with assessment environment may be particularly useful to students from families with less formal education and therefore more reliant on their teachers to help with test planning.

The results suggest that test takers prefer a range of physical movement/activities during testing, from no action to engaging in multiple movements/activities. Additionally, the results indicate that students have preferences for noises/sounds during testing, ranging from complete silence to loud noises. The results provide evidence that people have some combination of action and sound preferences during standardized testing. The defined rules common to many national standardized tests eliminate the abovementioned preferences and related conditions that students rely during exam preparation. Preferences for specific high-stakes testing conditions that mimic the environments that are generally adopted when studying and preparing for classroom tests have implications that are both proximal to and distal from the student. These proximal and distal levels of implications will be discussed in the following paragraphs below.

At the proximal level (i.e., the closest in proximity to the student), findings from this study may directly impact students and the parents/legal guardians of these students. Students and parents can be made aware of the relationship between the conditions they study under and the formal test environment. One example of study-testing environment incongruence is an “above average student” who studies every night, while drinking water and listening to music. In the school, his/her teacher allows students to have a bottle of water to drink during classroom quizzes and tests and plays background music while students complete assignments in class. Thus far, the student's study strategies have aligned with the in-school testing conditions and he or she has succeeded academically. However, when this student has to take a high-stakes exam with more controlled conditions, he/she performs poorly as the conditions that the student prefers and has come to rely on are not allowed. Therefore, the student's performance on the high-stakes test may not be reflective of his/her typical performance. With college admissions, scholarships, AP credit, and even high school graduation on the line, students have incentives to perform well on standardized tests. The TEOP can be administered to the student and the scores may be used to demonstrate any incongruence between student testing preferences and test conditions. Once this gap is demonstrated, the data and other supporting information can be used to tailor student studying and learning practices to mirror the environment of their upcoming standardized tests.

Modifications have been made to more accurately measure the performance of specific subgroups. Yet, there have been no substantive changes in how standardized tests are proctored to typical students in years, and it is unlikely that any changes are forthcoming. While major test environment adjustments are unlikely to come to fruition, there are other instances of parent groups effectively lobbying for educational change. For example, a parent-led campaign to adjust a grading scale was successful in Fairfax County, Virginia. Another instance of parent lead change occurred in Cooper City, Florida; where parent activism prevented the school board's planned attendance boundary changes (Public School Review, 2018). On a larger scale, parents in California used social media postings, paper flyers, mailing lists, and a support network of constituents to modify school lunch menus statewide (Public School Review, 2018). These examples, although dissimilar in context and cause, demonstrate how parents can mobilize and influence both local and statewide policy change.

Parents may have difficulty in effecting change to standardized testing procedures and rules at the state or national level. That is, even for the most organized and mobilized parents and parent groups, campaigning to transform the long-standing, high-stakes testing conditions may be futile without the support from state or national representatives in congress. With this limited range of influence, however, parents/legal guardians are still able to use the TEOP scores to inform the studying and test-taking strategies of their children. Results from this study indicate that a number of Millennials prefer testing conditions that are dissimilar to the standardized test center settings. These students may benefit from studying and preparing in environments that mimic those of the upcoming test. For example, from the student’s Action and Sound scores, parents/legal guardians can quickly determine if their child’s in-home exam preparation is aligned with the high-stakes testing procedures. The information from the TEOP preferences would be most beneficial to families with little or no experience with standardized testing and higher education planning. It would be a resource to help the parent/guardian better understand how test preparation often under the students preferred conditions and this can and often does differ from the defined environment prescribed by the major testing companies. This knowledge can allow parents to help the student study under surroundings congruent to the exam environment.

At the distal level (i.e., the farthest away in distance from the student), there are multiple groups of individuals with the potential to be impacted by the results from this study such as local, state, and national representatives and policy-makers, testing companies, K-12 schools/districts, and universities/colleges. Policy-makers, testing companies, and K-12 schools/districts and universities/colleges could (idealistically) use this measure to place test takers in a more comfortable environment to facilitate academic success. While it may be impossible for testing companies, schools, and universities to accommodate the infinite number of preferred testing conditions, the TEOP can be used in research studies to define a set amount of test environment "options" (in addition to the traditional format) for students scoring higher on the Action and/or Sound factors, such as testing rooms with music, testing individually, and testing outside.

The American Disabilities Act (ADA) is one example of policy-makers passing legislation to ensure that students with disabilities receive testing accommodations in order to accurately measure their "true" abilities and aptitude. The ADA obligates private, state, or local government entities offering licensing, certification, and credentialing exams in secondary, postsecondary, professional, or trade school to provide accommodations to persons with disabilities (ADA, 1990). However, if policy-makers are aware of a possible performance discrepancy for some typical students with dissimilar testing preferences that may be partially due to generational influences independent of the student's control, accommodations for all test takers may soon be a consideration.

The results suggest that students have individual "Action" and "Sound" preferences when testing that necessitate some level of customization. Increased customization of standardized test conditions would result in additional expenses, specifically to pay for more locations and proctors. These additional costs would either be passed on to the test takers, making the test more expensive and limiting participation, or cut into the testing company's profits. Furthermore, new testing options would require increased planning by both testing companies and individual test centers to properly organize the additional test environment possibilities.

Results from the current study indicated that most students had unique standardized testing experiences that were either helped or hindered by their preferences (or lack thereof) for specific environmental conditions. These findings have implications for K-12 schools/districts, namely the teachers and administrators employed by these schools/districts, that are increasingly asked to consider each individual student's needs and preferences prior to any school or district level educational planning. Scores from the TEOP may be used to create focused, educational interventions unique to each learner. Additionally, teachers, administrators, and especially academic guidance counselors, may consider using the TEOP results to develop both study and testing strategies that complement the learner's specific "Action" and "Sound" preferences.

The preliminary outcomes from this study could be useful to provide stakeholders' better information on their learners preferred testing conditions, which may increase the likelihood of achieving optimal performance. It is possible that students whose preferred environment is similar to the model currently used receive an inadvertent advantage over test takers whose favored environment is considerably different, resulting in obfuscation of their "true" abilities and aptitude. These traditional rules may not promote optimal standardized test performance in the current and upcoming generations of Millennial test takers who differ from their predecessors by having never known a world without constant auditory and sensory stimulation and

Extrinsic stimuli may influence performance regardless of the age group (Rasul & Bukhsh, 2011), but in the current study with Millennials, the cohort may be disproportionately impacted due to their unique upbringing. The additional impact of extrinsic stimulation on Millennials may necessitate some consideration when interpreting test scores compared to prior generations. Millennials' constant connectivity with technology (Oblinger, 2008), may contribute to the preferred testing conditions for optimal performance demonstrated in this study. Thus, it is important to develop additional current, valid, and reliable measures of self-perceived optimal performance in order to put students in the best position to succeed academically and in accordance with high stakes testing.

Limitations (Pilot Phases I and II) and Future Directions

Two sections of limitations are presented below: (1) Methodological/Statistical and (2) Psychometric. These limitations are accompanied by suggestions for future research.

Methodological/Statistical

The voluntary response sample is a limitation for Pilot Phase I, as there is no way to corroborate self-reported information. Unfortunately, this is often the nature of Internet-based survey research (Kline, 2013). Self-reported information can be problematic because subjects may forget details, exaggerate, or answer based on perceived social desirability (Northrup, 1996). Additionally, Phase I's sample may be disproportionately individuals who use social media, as the recruitment methods included an online invitation to participate in the study. Future research may consider using alternative recruitment strategies beyond social media and the Internet. In the main study (see Part II below), this limitation will be addressed by recruiting future participants' at their high schools in person, thus not limiting the study to only those who are active on social media or technologically savvy. Additionally, in the main study, all independent and dependent variable data will be de-identified, existing records (e.g., official transcript information) obtained with permission directly from each high school.

Given the geographic region from which the study participants were recruited and the disproportionately Caucasian composition (81.0%) of the Pilot Phase I sample (and 100.0% of the Phase II sample), the results may not be broadly generalizable (Dimitrov, 2010). These relatively homogeneous samples are in contrast to the demographic profile of college students provided by the National Center from Education Statistics (NCES). In the U.S., college student populations are approximately 58.0% Caucasian (NCES, 2017). Future studies should attempt to recruit a sample that is more representative of the population.

The Pilot Phases of this study used an informal mixed-methods design (i.e., using both quantitative and qualitative data; Subedi, 2016). One mixed-methods design that is frequently used includes conducting quantitative research followed by a qualitative phase (Ivankova, Creswell, & Stick, 2006). The rationale for this sequence is that the quantitative data and analyses provide a broad understanding of the research problems, and the qualitative data and analyses refine this understanding by exploring participants' attitudes and opinions in more detail (Ivankova, Creswell, & Stick, 2006). This design is limited by lengthy time requirements and resources necessary to collect and analyze both data types. The main study will only use quantitative research methods and analyses.

This study used quantitative methods and analyses (i.e., cross-validation) to explore and confirm the TEOP factor structure followed by a qualitative interpretive design. Because in qualitative research the data are derived from the meaning the participants give to the construct(s) of interest (Bogdan & Biklen, 1998), the researcher should have no pre-existing knowledge of results. In the current study, the quantitative phase was conducted first rendering a factor structure for test environment preferences. The subsequent qualitative phase included interview questions that may have been partially influenced by the results from the quantitative phase. It is, therefore, no coincidence that after the quantitative phase, corroborating themes emerged from the qualitative data. In the current study, the qualitative data and derived themes provided depth and support to the quantitative findings. Future research using a mixed-methods design may consider simultaneously analyzing the quantitative and qualitative data to limit the potential to create biased qualitative interview questions.

The Model Estimation method used in the CFA was Maximum Likelihood (ML). ML assumes that the data are normally distributed, but can support some deviations from normality (Hu, Bentler, & Kano, 1992). As five of the 11 original items were shown to be non-normally distributed (i.e., positively or negatively skewed), future research may consider model estimation methods other than ML such as Generalized Least Squares (GLS) and Unweighted Least Squares (ULS; Schumaker & Lomax, 2016).

The participants in this study were all at least six months, and in some cases several years, removed from taking the ACT or other standardized tests. Although slightly dissimilar from the traditional definition of history and maturation as threats to internal validity, variations of these threats can be applied in the context of this study. A historical threat to internal validity can be considered if an impactful event occurred between the date on which the student took the standardized test and the date that he/she completed the TEOP (Dimitrov, 2010). For example, the participant may have done better or worse on the ACT than anticipated. This knowledge of their score could impact the lenses through which they view to the testing environment in either a positive or negative direction. A maturation threat would confound results if physical, mental, or social development occurred between the taking of the ACT and completing the TEOP (Dimitrov, 2010). An illustration of this would be as the particpants becomes more aware of what works and does not work for them academically their preferences may change along with this new insight, thus their TEOP responses may differ from what they would have been at the time they sat for the ACT. In future research, participants in the main study will complete the TEOP concurrently (i.e., within the same two-week window) with the ACT in the Spring of their Junior year of high school.

The participants in this study were all current university students. Thus, it is surmised that these students obtained at least the minimum ACT necessary for admission to the university. Thus, a potential ceiling effect may exist, in which there are more scores at the upper range or level of the instrument (Wiersma & Jurs, 2009). This study's sample is limited by not recruiting and measuring students who performed poorly on the ACT, and therefore, may have been impacted by unfavorable conditions incongruent with their testing preferences. Future research should recruit a sample from the population with a representative distribution of ACT scores. Additionally, the main study will administer the TEOP and collect data on high school students with a wider range of abilities. As not all participants in the main study will be college-ready or meet admissions requirements, the main study will therefore be more representative of the population at large.

Psychometric

Self-report measures or scales that require individuals to inform on aspects of their own personality, emotions, cognitions, or behaviors, can be problematic (Kazdin, 2003). Although there are practical benefits to using self-report measures, this method of assessment enhances respondents' social desirability. The possibility of bias and distortion subjects involving their motives, self-interest, or to "look good" is elevated with a self-report measure. This limitation of self-reporting results that shines the participant in the best light is referred to as the halo effect (Wiersma & Jurs, 2009). The halo effect is related to the provocativeness of the measure. The constructs of interest measured in the TEOP are not controversial and students' preference would not be indicative of status. Although the directions indicated that the results would not be shared with anyone specifically, and they would be reported collectively, participants may still have fretted over others' perceptions. Future studies may consider deviating from online survey research methods and perhaps use direct observation or other measures.

Three items in the EFA portion of Pilot Study Phase I had factor loadings only minimally above the .32 threshold recommended by Tabachnick and Fiddell (2001) to account for the appropriate variance. Item 5 (i.e., I prefer to have water/beverage available) had a loading of .325 on the "Action" factor, Item 7 (i.e., I prefer to chew gum/candy) had a loading of .347 on the "Action" factor, and Item 2 (i.e., I prefer background/environmental noises) had a loading of -.447 on the "Sound" factor. Due to these low loadings a cognitive debriefing could have been considered to enhance the study. Cognitive debriefing is a method in which a measure is purposely assessed by a representative sample of the studied population to ascertain if the participants understand the items on the measure and are asking the appropriate questions to produce applicable answers (Clark K. , 2015). Due to both time and monetary constraints cognitive debriefing was not used in this research but could be considered for future studies with more resources.

Another psychometric limitation involves the sample used in the quantitative portion compared to the qualitative portion of the pilot studies. These samples may have markedly different perceptions and attitudes towards standardized tests. The high school students to be used in the main study are compelled by law to attend school and in most cases are required to pass a high-stakes examination to graduate. College students that participated in Pilot Phases I and II are under no obligation to attend school and graduation is achieved through accumulation of credits, not the passage of a high-stakes examination (i.e., certain career fields require passage of a high-stakes test to obtain a license). Therefore, high school students may view high-stakes assessments from the lens of something they must do well on to succeed in the future as compared to college students many of whom look at standardized tests as something related to their past. On the other hand, many, if not the majority of, university students will never be required to take a standardized test again. For these participants, standardized tests are distant memories compared to the immediate impact of this test on current high school students. Future research may consider studying a high school population that is more closely and directly impacted by standardized testing. Participants in the main study in Part II were recruited from a high school student population who completed the TEOP within two weeks of completing the official ACT examination.

Summary

The goal of the Pilot Study was to examine the test environment preferences amongst current university students. The current Pilot Study aimed to add to the limited body of theoretical and empirical literature examining standardized testing environmental conditions and the test takers conditional partialities. Furthermore, Phase I used cross-validation methods as a basis for testing the EFA and CFA models. Phase II used a qualitative interpretive design to provide further evidence to support the “Action” and “Sound” test environment affinities.

The main study in Part II will advance the outcomes from the pilot study by providing additional validation evidence for the TEOP scores in a population of high school students. The data will be analyzed using CFA to examine the psychometric properties of the TEOP in a new population. Hierarchical Multiple Linear Regressions will be conducted to provide evidence of the hypothesized relationships (i.e., Criterion-Related Validity – Concurrent) between the TEOP factors “Action” and “Sound” and aptitude outcomes (i.e., ACT scores).

PART II

CHAPTER I: INTRODUCTION (MAIN STUDY)

The pilot study in Part I used cross-validation techniques to address the psychometric properties of the newly created measure called the Test Environment for Optimal Performance (TEOP) in a college student population. This was supported by a second qualitative phase analyzing recent high school graduates perceptions of standardized testing and test environments. Part I provided evidence of the reliability and validity of the TEOP scores and that university students have action and sound preferences for environmental testing conditions.

The Limitations of the pilot study have been conferred in Part I. Among the Limitations that will be directly addressed are the amount of time between taking the standardized test and completing the TEOP. In Part II, participants took the ACT and TEOP within the same two week duration. Additionally, data collection methods were improved. The participants took the TEOP at their respective high schools and all academic records were provided by the school. The pilot study contained a college student population primarily in the Millennial Generation with a few participants from Generation Z (i.e., “Gen Zers” or “Gen Z”). As validity and reliability are contextual, the psychometric properties of a newly-developed measure/scale must be examined anew when using different populations and/or environments (Allen & Yen, 2002). Thus, Part II used a high school student population that is exclusively Generation Z.

High school students differ from college undergraduates in numerous ways. In the current context, the population is a member of a different generational cohort, specifically

Generation Z. Attending high school, unlike college, is required by law. Comparatively, university students can design their schedules around their own personal preferences such as enrolling in all evening classes (i.e., compared to the morning). High school class times and program offerings are set by the school district. State laws dictate basic high school graduation requirements independent of student interests. College students can design a schedule around their specific time preferences and preferred content areas. The majority of college students will never take a high stakes standardized test again (National Science Foundation, 2015). Students were administered the ACT to gain admission into college and, in some cases, to graduate from high school. This is contrasted with high school students, for whom standardized tests are one of many steps towards a successful academic and/or professional future.

Just as there are discrepancies between high school and college students, Generation Z is different from the Millennial Generation. Generation Z is the first cohort to have been raised entirely with immediate access to internet technology (Turner, 2015). The expectation of instantaneous information is a defining characteristic of this cohort (Williams, 2015). The technology revolution provided Gen Zers with opportunities for both mobility of and immediate access to communication, knowledge, and entertainment consumption (Jaconi, 2017). Generation Z spends less time with in-person social interaction than previous generations (Twenge, 2017). This potentially results in a group of young individuals who all have unique preferences.

The Generation Z cohort's ability and willingness to access information from anywhere at any time have also shaped their educational preferences. As instant information has become pervasive, evidence exists for the emergence of a new "social amnesia" phenomenon known as the "Google Effect." The Google Effect is the inclination to forget information that is readily available online (Sparrow, Liu, & Wegner, 2011). That is, the mobile device is potentially replacing the mind as a memory storage vessel. This will be detailed more in the Literature Review section below.

The main study (i.e., Part II) served as an extension of the Test Environment for Optimal Performance (TEOP) Pilot Study, to develop a valid and reliable measure of the same construct for current high school students' testing preferences. Part II will examine the psychometric properties of the TEOP in a sample of high school students, and investigate the relationship between TEOP scores and high-stakes standardized test scores. The main research questions (RQs) include: (1) "What are the psychometric properties (i.e., construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a high school student population?," and (2) "What is the relationship between the TEOP factor scores (i.e., Action and Sound) and high-stakes aptitude test scores (i.e., the ACT) in a high school student population?"

Statement of Problem

As discussed previously, there have been profound cultural and technological changes that have impacted this generation of high school students born after 1997. At the same time, the environmental conditions high stakes standardized tests are proctored under have remained stagnant. There is a lack of research on test takers preferred conditions for testing environments. Therefore, the psychometric properties of the TEOP and relationships between test environment preferences and standardized test aptitude will be examined.

Organization of Part II: Main Study

The organization of Part II (Main Study) will include a Literature Review (Chapter II) in addition to the review provided in Part I. The new sections included Generation Z, Generation Z and Learning, and the Google Effect. The Methodology (Chapter III) will include a Confirmatory Factor Analysis (CFA) examining the psychometric properties of the Test Environment for Optimal Performance (TEOP) in a high school student sample. Finally, a series of Hierarchical Multiple Regressions will be used to provide evidence of the relationships between the TEOP factors and aptitude outcomes. Following the Methodology, the Results (Chapter IV) from the Main Study will be reported. Finally, in the Discussion (Chapter V), results will be interpreted, which will involve implications, limitations, and future directions.

CHAPTER II: LITERATURE REVIEW

This section reviews the body of literature exploring Generation Z high school students' test environment preferences. In this literature review, research defining Generation Z is outlined followed by existing research on this generation's learning styles and preferences. The section concludes with a summary of "the Google Effect."

Generation Z

Generation Z is a common name for the first post-Millennial generation. Generation Z is the youngest of five active generations. Members of Generation Z were born between approximately 1997 and 2012-2015. Researchers have not yet established a final generational year, and will determine it as the group continues to develop (The Center for Generational Kinetics, 2017). Generation Z consists of 65 million members in the United States alone and comprises approximately 26% of the population (Claveria, 2018; Kalkhurst, 2018). There are several characteristics of Generation Z that distinguishes it from preceding cohorts. Generation Z is more demographically diverse than prior generations. Members were born to older mothers, into smaller families, having the fewest siblings of any preceding generation (McGrindle Research, 2010). Generation Zers are labeled as "instant minded" with a faster life rhythm, tending to be more distinctive, and self-directed (Ferinez, Hortovanyi, Szabo, & Tarody, 2010). This generation has expanded upon the multitasking, instant gratification, and personalized lifestyles that began to form with Millennials.

Members of Generation Z are categorized as highly connected, with active use of communications and media technologies. Generation Zers have been branded with superficial and divided attention in part because they simultaneously use multiple kinds of media (Tari, 2010). While Millennials average three screens per day, Generation Z uses five – a smartphone, laptop, television, desktop, and tablet (Kalkhurst, 2018). This cohort has been labeled multitaskers, due to their penchant for instantaneously listening to music, studying, writing an e-mail, instant messaging, all while arranging to meet with friends (Pal, 2013). Generation Zers compartmentalize their social media time. Communally, they share certain types of content on particular social channels. On Snapchat, they share real-life moments. And potentially the most damning characteristic, this generation obtains their "news of the day" from Twitter and Facebook.

All of these activities are completed in quick interactions, lasting only a few seconds each (Response Media, 2017). According to Tabscott (2009), Generation Z has eight special characteristics – freedom, customization, collaboration, scrutiny, integrity, fun, speed, and innovation (Tapscott, 2009). Members of Generation Z, more so than prior generations, value self-determination and freedom of choice, desiring to personalize things, and to make them their own (Tapscott, 2009). This cohort is used to technology that automatically adjusts to their preferences and expects everything to be personlized (Lambert, 2018). Generation Zers very much desire what they want, how they want it, and when they want it, which also includes their education and how they learn.

Generation Z and Learning

A major difference between Generation Z and previous generations is the variety of alternate and progressive forms of education available that did not exist even a decade ago (Reynolds, 2018). The cultural and technological changes in society and communication have impacted the learning styles and practices of Generation Z. Differentiated instruction and increased learning platforms, current students reject being passive learners, and expect to be fully involved as part of the learning process themselves. Formal learning in schools is structured by subject, while Generation Zers are living in a hyperlinked world. Educators convey formal instructions, however learners are experiential and participative (McCrindle, 2016). Academic knowledge and memory is assessed in standardized examinations, yet Generation Z lives in an "open-book universe" that is only a few clicks away from the entirety of world knowledge.

Gen Z students tend to thrive when they are involved in the learning experience (Barnes and Noble College, 2015). They have the ability to find necessary information and teach themselves new skills without the help of intermediaries. Gen Z members are multimodal learners, preferring to learn through a variety of learning channels, and multiple ways of informational delivery. Evidence suggests that frequent social media users pay better attention to digital environments and have lower attention scores in other environments (Microsoft Corporation, 2015). Sixty-seven percent of Canadian 18-24 year olds ($N = 2{,}000$) have a difficult time concentrating enough to stay on task, with 61% of them getting distracted by unrelated thoughts or daydreams (Microsoft Corporation, 2015). This concentration and distraction can be attributed to Generation Z members having only known a world that brings instant change and/or stimulation.

Generation Zers typically prefer video content over the written word, with 85% ($N = 1{,}004$) of those surveyed reporting that they watched an online video to learn a new skill in the past week (The Center for Generational Kinetics, 2017). Seventy four percent

(*N* = 1,004) of Gen Zers spend two or more hours a day on social media (The Center for Generational Kinetics, 2017). Additionally, 40% of Generation Z (N = 1,004) prefers to have working WiFi over working bathrooms. The use of surveys, games, and neurological research, to study attention spans found that people tend to lose concentration after eight seconds. This is down from 12 seconds in the year 2000, and provides evidence of the impact of digitalization on the brain (Microsoft Corporation, 2015). Seventy-six percent (N = 2,000) of the cohort considers multitasking as the only way to get all of their required tasks done (Microsoft Corporation, 2015). The ability to multitask has considerably improved over this 15-year span, with frequent technology users found to be “better multitaskers” (Microsoft Corporation, 2015). Generation Zers see multitasking/task-switching as a way of life.

Researcher Jean Twenge prefers the term IGen over Generation Z because 2/3 of teenagers (*N* > 5,000) own an iPhone (Twenge, 2017). The number of high school seniors who read a non-required book or magazine nearly every day dropped from 60 percent in 1980 to only 16 percent in 2015. Related to this, the average SAT critical reading scores have dropped 14 points since 2005 (Twenge, 2017). Circumstantial evidence suggests that internet searches and the informality of text message communications leads to more efficient reading at the expense of longer detailed passages (Wolf, 2008). Reading is not a human instinct like, for example, breathing, If the way people read changes books and newspapers to short passages on the internet, their minds will adapt (Wolf, 2008). Research suggests that individuals in 2018 are not reading less than in the past, but rather reading differently. One study found that when reading online people tend to “hop” from one site to the next, seldom reading more than a page or two, in contrast to prior generations who read long passages in one sitting (Rowlands et al., 2008). Anecdotally, college professors say that students have more trouble reading longer text passages than previous generations, and rarely read the required textbook (Twenge, 2017). This generational change in reading habits undoubtedly has an impact on academic performance.

“The Google Effect”

Generation Z, the current cohort of high school students has spent the entirety of their lives with access to search engines. The ability to hunt for information from anywhere at any time has impacted how these students learn and retain information (The Kapersky Lab, 2015). Google is the preeminent search engine in the United States with 65.6% of all searches and as of May 2018, the most used website on the planet (Cox, 2017; Lipman, 2010). Furthermore 21 of the 67 most frequented websites are owned by Google (Amazon.com, 2018). The number of daily Google searchers has grown from 9,800 in 1998 to in excess of 3.5 billion in 2017 (Mangles, 2018). The increase of immediate access to knowledge is evidenced by mobile phones accounting for over half (50.3%) of all searches in 2017, up from less than 1% (0.7%) in 2009 (Mangles, 2018). One study states that search engines impact the way people think by providing an illusion of accessibility, implying that the provided links are “stopping points,” and masking the interconnectedness of information through the appearance of linearity (Heick, 2018). The ability to simply “Google” any problem suggests to users that all answers are found easily by asking a simple question. Immediate websites provided by search engines suggests to the user that they have reached a stopping point and no further inquiry is required, while simultanesouly concealing possible connections with related knowledge.

The advent of Google and other search engines has impacted human memory and the ability to recall information. A surge of digital amnesia or “The Google Effect” has “sprung up” along with immediate access to information becoming commonplace (The Kapersky Lab, 2015). The Google Effect is the tendency to forget information that can be easily accessed online by using search engines like Google (The Kapersky Lab, 2015). Recent studies submit that when faced with difficult inquiries, people are inclined to think about technology and expect to have future access to information, and there are lesser rates of recall of the information itself and increased recollection instead for where to access it (Sparrow, Liu, & Wegner, 2011). The Internet has become a primary method of recall.

As the opportunity to instantaneously access knowledge over the last decade has become the norm, people show signs of withdrawal when they cannot immediately look up the answer to a question (Sparrow, Liu, & Wegner, 2011). Individuals tend to forget information that they believe will be available to them later via a search engine and remember things they think will not be obtainable later. This is known as transformative memory and is evidence that humans are adapting to computers and communication technology. People are using the internet to replace their memory.

Human beings are becoming symbiotic with technology and recollect not the knowledge itself, but rather, where to access it (Sparrow, Liu, & Wegner, 2011). This essentially means that individuals, instead of using their brains to store knowledge over time, are relying on Google and other search engines as their memory bank (Nielsen Company, 2012). Simply put, the brain adapts to meet needs and memory storage is becoming less of a necessity to survival (Mastin, 2018). The more people use Google and other search engines, the less likely they retain what they "googled" (Small, Moody, Siddarth, & Brookheimer, 2009). Findings reveal that the Google Effect extends beyond online facts and includes important personal information such as telephone numbers and addresses (The Kapersky Lab, 2015). Neuroimaging of frequent internet users shows there is double the activity in the short term memory as intermittent users when online (Sparrow, Liu, & Wegner, 2011). The brain has begun to disregard information that can be found online, and the disconnection becomes stronger with more internet encounters. Therefore, it is reasonable to surmise that the Google Effect and the evolution of how humans retain information may also result in changing preferences for testing conditions.

CHAPTER III: METHODOLOGY

The main study extends the findings from the pilot study by providing more validation evidence for the scores on the Test Environment for Optimal Performance (TEOP) in a population of high school students. The data will be analyzed using Confirmatory Factor Analysis (CFA) to examine the psychometric properties of the TEOP in a new population. Hierarchical Multiple Regressions will be used to provide evidence of relationships between the TEOP factors “Action” and “Sound” and aptitude outcomes. The sections in Chapter III (Methodology) include purpose and research questions, participants, measures, procedures, and data analysis.

Purpose and Research Question: Main Study

With the explosion of technology and social media since the 2007 release of the iPhone, society and education have transformed substantially. This electronic communication eruption, coupled with the 2003 No Child Left Behind (NCLB) Act and subsequent legislation, have substantially increased the profile of high-stakes assessment in education. At the same time, there has been no change in the rules and procedures for proctoring and taking high-stakes standardized exams. It is important to build on the Test Environment for Optimal Performance (TEOP) Pilot study to develop valid and reliable measures for high school students, of self-reported optimal performance conditions in order to give stakeholders a method with which to assess current students’ (i.e., Gen Z) testing preferences when interpreting high-stakes standardized assessment scores. Additionally TEOP results can allow for comparisons of ACT outcomes between those whose preferences align with current testing practices and those whose differ.

A valid and reliable measure of test environment preferences is arguably of more value for high school students than college students. The vast majority of undergraduate students will not attend graduate school, and therefore will never take a high-stakes standardized test again. High School students on the other hand will take standardized tests like the ACT or SAT to get into college, many will take AP exams while trying to earn college credit in high school, a large number of states require standardized tests as a graduation requirement, and some school districts use common end of course exams as a means to measure student success. A typical US student takes 112 mandated standardized tests between preschool and high school graduation (Hart, Casserly, Palacios, Corcoran, & Spurgeon, 2015). Additionally the greater control exerted by parents and teachers over juvenile high school students as compared to adult university students makes this an ideal time to measure the students preferred testing conditions. This knowledge will allow for in school interventions and the development of better study habits.

The TEOP, as described in the Part I, is a short (9-item) measure used to assess students' personal preferences for particular environmental conditions while taking a test. Each statement in the TEOP includes a preference for either noise/sound or physical movement while taking a test. Students can respond on a 5-point Likert scale to what degree the statement describes their test-taking tendencies (i.e., from "Never Me" to "Always Me"). Confirmatory Factor Analysis (CFA) will be used to address the two main research questions and provide evidence of internal consistency reliability for each factor. Hierarchical Multiple Regressions will be used to examine if there are relationships between the TEOP factors and aptitude outcomes.

The purpose of this study was to investigate the psychometric properties of a newly-developed measure entitled the Test Environment for Optimal Performance (TEOP) in a sample of high school students, and to explore the relationship between TEOP scores and high-stakes standardized test scores in a sample of high school students. The main research questions (RQs) include: (1) “What are the psychometric properties (i.e., construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a high school student population?”, and (2) “What is the relationship between the TEOP factor scores (i.e., Action and Sound) and high-stakes aptitude test scores (i.e., the ACT) in a high school student population?” Research Question One was an extension of the pilot study. This will quantitatively validate the TEOP with a high school student sample, specifically high school juniors. Research Question Two explores the relationships between “Action” and “Sound” preferences and external measures of interest. Additionally it provides more psychometric evidence by addressing concurrent validity.

As more information about the psychometric properties of this measure was explored, the main focus in conducting this study (as an extension of the Pilot) was to address validity. The process of construct validation is continuous, and therefore, more construct validity evidence is needed when using an existing measure with a new population or in a different context (Clark & Watson, 1995). Except for some replication studies, no two are exactly the same in terms of the sample, context, and the inclusion/exclusion of other variables. Every study should provide evidence of a measure’s psychometric properties, as documenting any differences in the reported reliability and validity of the measure's scores can contribute to construct development and its associated theoretical nomological network (Schreiber, Stage, King, Nora, & Barlow, 2006). Thus, in the current main study, criterion-related validity was addressed by examining the relationship between the TEOP scores and high-stakes standardized test performance, extending the use(s) and purpose(s) of the measure.

Context: Main Study

In the current main study, the population of interest included high school students who are categorically "Generation Z," and raised in a time of perpetual technological change (Neal, 2018). While this cohort has faced similar standardized testing demands as the preceding generation of "Millennials," their unique upbringing makes them somewhat different with regards to their testing needs. The current study's population includes 11th grade (Junior) high school students who were born between 2001 and 2002. The attacks of September 11th are only a "story" in their history books, much like Pearl Harbor is to "Millennials." They differ from previous generations in that "Generation Z" has never been disonnected from technology and the internet travels with them (Neal, 2018). These particiapants have been able to "stream" content (i.e., watch live videos or movies on wireless devices) their entire lives, with all of the world's knowledge and entertainment on demand any time of day. Additionally, one of the most defining characteristics of this cohort includes the prevalance of social media and social-networking site (SNS) use. Theses high school students have grown up alongside social media and it is natural for them to document their lives on Snapchat and Instagram (Neal, 2018).

Aside from this generation's obvious use (and near abuse) of social media, these high school juniors were educated from a young age in schools that tailored instruction to their individual abilities and interests. Their teachers used differentiated instruction to facilitate educational entry points, learning tasks, outcomes, and assessments structured to each students' specific needs (Subban, 2006). This generation has been conditioned to expect a much greater variety of product and service selectivity. They have grown up with a huge assortment of choices and the expectation that such abundance is their birthright.

As media content options and entertainment delivery mechanisms have increased with the advent of portable technology high school student reading behaviors have changed. The percentage of high school seniors who read a non-required book or magazine nearly every day dropped from 60 percent in 1980 to only 16 percent in 2015 (Twenge, 2017). Twenge (2017) reports that college professors say that students have difficulty reading longer text passages, and seldom read the required textbook. Simultaneously, average SAT critical reading scores have dropped 14 points since 2005. This decrease in reading for pleasure and the coinciding drop in standardized reading scores suggest that this "Generation Z" study habits and testing preferences may not align as well as they did for prior generations. As schools prepare their students for academic success both in the classroom and in standardized testing it is crucial to be cognizant that the individual preferences and behavioral inclinations of the learners may be different than those who preceded them.

Participants and Procedure

A total of 402 high school 11th grade students completed the TEOP. Fifty one cases were removed due to incomplete or inconsistent information provided. The removal of these cases and description of the final analysis sample will be presented in detail in the Results section (Chapter IV) below.

The measure was uploaded to a survey-hosting website addressing participant demographics and the nine items of the TEOP. High school juniors were recruited to participate via direct contact at their public and private high schools. A short recruitment message was used, "Your participation in this short survey study can help identify student preferences for optimal test performance, which can impact best practices for higher education and the workforce. Your responses will be completely anonymous, and no identifying information will be collected. This study has been approved by Kent State University Institutional Review Board (IRB # 18-114). The online survey for each participating school was open for one day before it was closed at the end of the study.

Measures

The TEOP is a measure that contains nine items using a 5-point Likert response format to measure the preferred environmental conditions for optimal performance while testing. On the TEOP Likert scale (i.e., "Never Me," "Rarely Me," "Sometimes Me," "Often Me," and "Always Me"), participants rated how strongly a particular environmental testing condition was indicative of their personal preference. Environmental testing conditions included preference for silence, music, water, food, and the ability to walk around.

Data Analysis

The first research question addressed in Part II is, “What are the psychometric properties (i.e., construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a high school student population?” In the Pilot Phase in Part I, university student data were analyzed using cross-validation techniques (i.e., EFA and CFA) to demonstrate psychometric support. Based on the factor structure in Part I, a second qualitative phase was used to provide evidence that students have both “Action” and “Sound” preferences when taking standardized tests.

In Part II, or the main study, only CFA was used to confirm the factor structure of the TEOP in a high school student population. It is important to confirm the factor structure in a high school population because validity is contextual, and it cannot be presumed that a measure validated on a university sample and applied to a high school sample will yield the same results.

After the TEOP was confirmed in a high school student sample, researchers, high school administrators, and teachers can safely and reliably use it with their students to establish if their students have “Action” and or “Sound” preferences when testing. Once the TEOP has been validated on a high school student sample, the results can be used to provide targeted interventions to high school students prior to them taking high-stakes standardized tests. For example, these interventions might be able to help students who have a strong preference for “Action” to self-regulate/monitor their behavior while studying or participating in test preparation. During lower stakes classroom exams or practice tests these high “Action” preference students can condition themselves to limit any behaviors that might negatively impacting their test performance.

Prior to conducting CFA, four multivariate assumptions must be met. These assumptions include: (1) Normality, (2) Sufficient Sample Size (3) the correct a priori model specification, and (4) Data must be drawn a random sample. Normality is observed when the distribution of each observed variable (i.e., the TEOP items) is normal. Pett, Lackey, and Sullivan recommend 10 subjects per item as sufficient to meet the sample size assumption (Pett, Lackey, & Sullivan, 2003). While the rule of thumb for appropriate sample size is 10 per item, to be more precise a power analysis was conducted, and it was determined that 69 participants were necessary to meet this assumption. The correct a priori model specification is met when the theory based factor structure used is accurate. The assumption of a random sample is met if the subjects being measured were randomly drawn from the population (Lomax & Hahs-Vaughn, 2012).

Model specification

Specifying the theoretical model is conducted based on theory and previous inquiry (Schumaker & Lomax, 2016). This stage consists of specification of the number of factors within the data, the factors that are associated with the observed variables, the factors that are expected to correlate, the errors that are expected to correlate, and the factor loadings that should be held equal (Dimitrov, 2010).

Model identification

Model identification establishes if the hypothesized model is identified. The model is identified when there is sufficient information available to make the parameter judgements. The number of free parameters analyzed must be less than or equal to the number of distinct values in the matrix *S* in the model. The distinct values in the matrix *S* can be determined using the formula noted in Part I (p[p+1]/2). If the number of distinct variables in the sample matrix *S* is greater than or equal to the number of free parameters the model is considered identified (Schumaker & Lomax, 2016).

Model estimation

Model estimation involves subtracting the elements in the population matrix from the components in the sample matrix to examine if the χ^2 value will equal zero (i.e., a perfect model fit; Schumacker & Lomax, 2016). Maximum Likelihood (ML) estimation is recommended to produce factors that replicate the correlations (Thompson, 2010). This estimation method assumes that the data are normally distributed. Generalized Least Squares (GLS) and Unweighted Least Squares (ULS) are both less stringent against the normality assumption and will be considered in the event of nonnormal data (Schumaker & Lomax, 2016).

Model testing

Model testing defines how well the data fit the model and is examined through goodness-of-fit indices. The chi-square fit index is examined along with other goodness-of-fit indices such as the Goodness-of-Fit Index (GFI), the Adjusted Goodness-of-Fit Index (AGFI), the Standardized-Root-Mean-Square Residual (SRMR), and the Root-Mean-Square-Error of Approximation (RMSEA; Steiger, 1990).

Model modification

Model modification is used when the model does not fit the data. When a modification is added, such as deleted or adding a path, or including an error covariance, support for the modification should be included (Schumaker & Lomax, 2016). After these modifications are made, the model is tested again.

The second research question in Part II was, “What is the relationship between the TEOP factors (i.e., “Action” and “Sound”) and student standardized test scores (i.e., the ACT)?” Multiple Regression was the primary statistical procedure used to address this research question. The current study used Multiple Regression analyses to explore the relationship between testing environment preferences, as measured by the TEOP, and performance outcomes, as measured by standardized ACT test scores. Multiple Regression also permits for the use of numerous independent variables, which can be organized into meaningful groupings. These groups of IVs included demographic covariates, academic covariates, behavioral covariates, studying/testing covariates, and TEOP covariates.

Multiple Linear Regression (MR) is commonly used across social science disciplines, including Education, to describe the nature of relationships between a set of predictor variables and an outcome. This technique allows for multiple independent variables when testing models that are hypothesized to predict the dependent variable being researched. The term “Multiple Linear Regression,” defines both the predictive and outcome components of the model. First, the word “Multiple” refers to the Independent Variables (IVs). MR includes at least two input variables (i.e., IVs or “Predictors”) that are used in combination to predict or explain the output variable (i.e., Dependent Variable [DV] or "Outcome"). MR is an extension of Simple Linear Regression (SLR), which is used to explore the predictive relationship between one IV and one DV (Keith, 2006). That is, SLR and MR both have one DV; however, in MR there are multiple IVs (Stevens, 2009). MR is used to assess the size and significance of the effects of several independent variables on a dependent variable (Stevens, 2009). That is, MR uses a set of independent variables to predict one outcome variable.

Second, the word "Linear" defines the DV and the type of regression being conducted. MR is an umbrella term that encompasses a family of models with several types. The most common are linear, logistic, multinomial, and ordinal. These MR types have comparable structures with different DV measurement levels. For example, "Linear" suggests that the DV in this type of regression model is a continuous variable (i.e., interval or ratio level). "Logistic" designates a DV that is a categorical, specifically dichotomous, variable (i.e., nominal level). The remaining two types are categorized as "Polytomous," in that the outcome variables have more than two categories with the absence or presence of inherent ordering across the categories defining the type to use. "Multinomial" indicates a DV that is a categorical variable with more than two unordered categories (i.e., nominal level). "Ordinal" indicates a DV that is categorical with more than two ordered categories (i.e., ordinal level). In the current study, the main DVs for RQ 2 (i.e., students' scores on the ACT exam both by category and comprehensive) were continuous, approximately normally distributed variables (i.e., interval level). Thus, this study used Multiple Linear Regression as the dependent variable across all models will be interval-level variables.

Method of Entry

One further descriptor that can be used when describing the various types of regression models is the method used to enter the variables into the model (i.e., the "method of entry"). A variety of methods of entry exist (e.g., hierarchical, stepwise, forward, backward, free) and the current study used "sequential entry," which is more commonly known as "hierarchical entry." In hierarchical entry, the IVs are entered into the model according to a specific hierarchy or order. In hierarchical regression the independent variables are chosen based on prior research and theory, with the researcher determining the order the IV's are entered into the model (Field, 2014). Hierarchical regression is suitable when there are numerous independent variables and the researcher is interested finding the most influential subset. The principal purpose of hierarchical regression is explanation (Lomax & Hahs-Vaughn, 2012). Hierarchical regression allows the results to show the variance explained in the DV by a particular IV or grouping of IVs while controlling for the effects of the other IVs previously entered into the model (Keith, 2006).

MR with hierarchical entry (i.e., also called "Hierarchical Multiple Regression" or "Hierarchical Multiple Linear Regression") enters explanatory variables into the model in "blocks" (i.e., typically conceptually or methodologically related groups of IVs) with the order of these blocks determined by their theoretical importance to the outcome. In the current study, Hierarchical Multiple Linear Regressions were used to investigate the participants performance on the ACT (i.e., Comprehensive ACT Score, ACT Math Score, ACT Science Score, ACT English Score, and ACT Reading Score), related to four blocks of IVs. These blocks include the following information: (1) Demographic/Behavioral data, (2) Study/Testing Data, (3) High School Academic data, and (4) TEOP Data. The theoretical hierarchy of blocks was sequenced so as to go from proximal to distal. The first two blocks “Demographic/Behavioral” and “Study/Testing” are more personal to the students’ identity while the third and fourth blocks “High School Academic” and “TEOP” are more related to school day data. The statistical analyses addressed the extent to which the more subjective independent variables contribute to the explained variation in the dependent variables over and above the contribution of more objective variables.

The statistical analyses outlined above involve entering the groups of variables (i.e., the “blocks”) into the Hierarchical Multiple Regressions in a sequential order. As the blocks were included in the models, the change in explanatory power associated with the addition of each group of variables was observed. The statistical analyses addressed the extent to which variable groupings contribute to the explained variation in the dependent variables over and above the contribution of other independent variable blocks.

A four step Hierarchical Multiple Regression was conducted with the comprehensive ACT Score and the ACT scores of the four ACT subcategories (i.e., Math, Science, English, and Reading) as dependent variables. Demographic/Behavioral (i.e., Gender, Race, public or private school, Mother and Father’s education as proxies for Socioeconomic Status, extracurricular participation, and do they exercise,) data was entered at step one. Study/Testing data (i.e., Good test taker, difficulty staying focused when testing, and preferred testing time congruency) was entered at step two. The High School Academic data (i.e., Number of times taking ACT, Cumulative Grade Point Average (GPA), favorite subject and if they take AP/Honors courses) was entered at step three. Finally TEOP (i.e., TEOP score on "Action" and "Sound" factors) data was entered at step four. The multi-categorical or ordinal IVs with three or more groups will be dummy coded using the lowest level as the reference group. Table 13 (see below) summarizes the variables (i.e., IVs/Covariates and DVs) in the model by hierarchical block, level of measurement, and the coding applied to the variables.

Table 13

Main Study Demographics (*N* =*351*)

Variables	*M*(*SD*) or (%)	Regression	*M(SD)* or *n*(%)	Coding
Gender				
Female	182(51.9)	SAME	SAME	0
Male	169(48.1)	SAME	SAME	1
Ethnicity				
White/Caucasian	291(82.9)	Other	60(17.1)	0
Black/African American	29(8.3)	White	291(82.9)	1
American Indian/Alaskan Native	1(.3)			
Asian	5(1.4)			
Hispanic/Latino	11(3.1)			
Bi or Multiracial	14(4.0)			
High School Type				
Public	132(37.6)	Public	132(37.6)	0
Private	219(62.4)	Private	219(62.4)	1
Mother's Education				
Not Sure	19(5.4)			
No High School Diploma	17(4.8)			
High School Diploma	58(16.5)			
Some College	62(17.7)	< Bachelors	156(44.4)	0
Bachelor's	113(32.2)	≥ Bachelors	193(55.0)	1
Master's	60(17.1)			
Professional	8(2.3)			

Doctorate	12(3.4)			
Missing	2(.6)			
Father's Education				
Not Sure	32(9.1)			
No High School Diploma	14(4.0)			
High School Diploma	78(22.2)			
Some College	56(16.0)	< Bachelors	180(51.3)	0
Bachelor's	97(27.6)	≥ Bachelors	169(48.1)	1
Master's	44(12.5)			
Professional	14(4.0)			
Doctorate	14(4.0)			
Missing	2(.6)			
Number of ECAs				
0	65(18.5)	No ECAs	65(18.5)	0
1	66(18.8)	= 1 ECAs	66(18.8)	1
2	66(18.8)	= 2 ECAs	66(18.8)	2
3	53(15.0)	= 3 ECAs	53(15.0)	3
4	43(12.3)	= 4 ECAs	43(12.3)	4
5	26(7.4)	≥ 5 ECAs	58 (16.5)	5
6	14(4.0)			
7	9(2.6)			
8	1(0.3)			
9	3(.9)			
10	5(1.4)			
Exercise				
No	100(28.5)	SAME	SAME	0
Yes	250(71.2)	SAME	SAME	1
Missing	1(.3)	--		
Good Test Taker				
No	165(47.0)	SAME	SAME	0
Yes	184(52.4)	SAME	SAME	1
Missing	2(.6)	--	--	
Trouble Focusing when Testing				
No	146(41.6)	SAME	SAME	0
Yes	204(58.1)	SAME	SAME	1
Missing	1(.3)	--	--	
Test Time Congruency with ACT				
No	254(72.4)	SAME	SAME	0
Yes	97(27.6)	SAME	SAME	1
GPA (4.0 Scale)	3.10(.73)	--	--	--
ACT Attempts				
1	213(60.7)	= 1 Attempt	213 (60.7)	0
2	98(27.9)	> 1 Attempt	138(39.3)	1
3	34(9.7)			
4	5(1.4)			
5	0(0.0)			
6	1(.3)			
Taken an AP/Honors Class				

No	153(43.6)	No	153(43.6)	0
Yes	198(56.4)	Yes	198(56.4)	1
Favorite Subject				
Math	77(21.9)	SS+Art+PE	143(40.7)	0
Science	80(22.8)	MA+SC+En	206(58.7)	1
Social Studies	63(17.9)	g		
English	49(14.0)			
Art	44(12.5)			
Physical Education	36(10.3)			
Missing	2(.6)			
Action	3.48(.97)	--	--	--
Sound	2.32(.93)	--	--	--
Main Dependent Variables		--	--	--
ACT Composite	21.45(5.41)	--	--	--
ACT Math	21.18(5.24)	--	--	--
ACT Science	21.62(5.19)	--	--	--
ACT English	20.93(6.54)	--	--	--
ACT Read	21.95(6.40)	--	--	--

Note. ECA = Extra-Curricular Activities. SS = Social Studies. PE = Physical Education. MA = Math. SC = Science. Eng = English.

Preliminary Analyses

Before conducting the MRs, outliers were examined, and assumptions tested (Sevier, 1957). Outliers are not an issue as the TEOP is scored on a 5- point Likert Scale.

Extreme values (i.e., outliers) are defined as data that fall outside the normal range of values on one or more variables. Extreme values are more likely to produce a skewed variable in the data set, and can increase the risk of committing a Type 1 or Type II Error (Lomax & Hahs-Vaughn, 2012).Three typical steps will be taken to examine outliers: (1) Reviewing the data for transcription errors, (2) Coding the outlier cases so that they will not be treated as valid data, and (3) Considering outlier cases for deletion. If there are outliers present, the data will be reviewed to determine if they are a result of reporting error or truly an outlier. If an outlier is detected, but they are not the result of an error, the final decision will be to determine if the case(s) should be deleted to minimize any distortion in current study.

Prior to conducting the MRs, several critical assumptions were investigated to increase statistical validity (Keith, 2006). Consideration of the MR assumptions will improve the conclusions drawn that will contribute to the theory in the literature base. The assumptions of MR include Normality, Linearity, Homoscedasticity, Independence of Errors, and depending on the resource, Multicollinearity. If any of these assumptions are not met, Type I or Type II Errors may be committed, or an over- or under-estimation of the effect sizes may result (Lomax & Hahs-Vaughn, 2012) When assumptions are violated, the exactness of the inferences from the analyses and results are inaccurate. In the following sections below, each assumption will be defined with the corresponding consequences for violating each one.

Normality

Multiple Linear Regression assumes that variables are normally distributed (Lomax & Hahs-Vaughn, 2012). More specifically, the analysis requires that the errors between the observed and predicted values (i.e., the residuals of the regression) should be normally distributed (Lomax & Hahs-Vaughn, 2012). Hayes (2013) emphasizes that this assumption is one of the least important in regression. In other words, regression is considered robust if the normality assumption is violated slightly (Hayes, 2013). The assumption examines the errors to determine if they are normally distributed. Normality is analyzed by visual inspection of a graph to determine if the data are skewed. If the data are not skewed, the graph of the residual values will approximate a normal curve (Keith, 2006). This assumption can be checked by inspecting a Histogram or a Q-Q-Plot of the residuals. Normality can also be investigated with Goodness-of-Fit tests (e.g., the Kolmogorov-Smirnov test); however, these tests must be conducted on the residuals and not the raw data.

Linearity

Multiple Linear Regression necessitates that the relationship between the IVs and DV is linear. The dependent variable should have an approximately linear relationship with each of the independent variables, controlling for the other predictor variables in the model. Linearity is considered the most important assumption, and if violated, the application of a non-linear relationship to a linear model nullifies the entire analysis (Keith, 2006) If Linearity is not present, all the estimates of the regression coefficients, standard errors, and tests of statistical significance could be biased (Keith, 2006). Similar to Normality, if this assumption is not met the results will under- or over-estimate the true relationship between the model variables and increase the risk of committing Type I and Type II Errors (Lomax & Hahs-Vaughn, 2012). To investigate this assumption, scatterplots of the IV and DV residuals were reviewed. Scatterplots can graphically portray the presence or absence of a linear relationship.

Homoscedasticity

This assumption indicates that the variance of error terms is similar over the range of IVs in the regression model (Keith, 2006). That is, the data should be to some extent linear in nature and free of any patterns or clustering (Dimitrov, 2010), in other words the residuals should spread consistently across levels of the independent variables (Lomax &

Hahs-Vaughn, 2012). When a heteroscedastic pattern is present, results from the analysis will be biased. The biased results can weaken the estimate of the omnibus test and the strength of the individual predictors in the model leading to an increased risk of committing a Type I Error. Multiple resources note that MR is fairly robust to violations of this assumption (Keith, 2006, Lomax & Hahs-Vaughn, 2012). One technique to test the Homoscedasticity assumption is to examine a scatterplot of residuals against the predicted values.

Independence of Errors

The assumption of Independence of Errors states that the estimation errors are statistically independent. This means that the residuals should be uncorrelated. Specifically, the Independence assumption is violated if the data are not sampled/collected independently of the population (Dimitrov, 2010). One strategy to address this assumption is to determine if the data contain repeated measurements, as longitudinal data with data collection at multiple time points is an obvious violation of this assumption. The Independence of Errors indicates that cases in the study responded independently of each other (Stevens, 2009).

Multicollinearity

MR assumes that the IVs are not extremely correlated with each other. Multicollinearity (also called Collinearity) refers to the undesirable condition that the IVs are highly correlated (Keith, 2006).This occurs when some IVs are highly associated with each other, or when an IV is a linear combination of the other IVs. As the IVs increase in their similarity, it becomes more difficult to separate the effects of the variables on the outcome. Thus, it is desirable that the IVs are more highly correlated with the DV than with each other (Keith, 2006). Multicollinearity is present when there are moderate to high correlations between two or more IVs (Dimitrov, 2010) This can be checked by examining the tolerance and variance inflation factors (VIF) and/or a correlation matrix. A tolerance value below 0.1 or a VIF above 10 is cause for concern and may indicate Multicollinearity (Keith, 2006).When examining the correlation matrix, the magnitude of the correlation coefficients should be less than .80.

CHAPTER IV: RESULTS

Research findings presented in this chapter consist of a Confirmatory Factor Analysis (CFA) to answer the first research question "What are the psychometric properties (i.e., construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a high school student population?" as well as a Hierarchical Multiple Regressions to answer the second research question "What is the relationship between the TEOP factor scores (i.e., Action and Sound) and high-stakes aptitude test scores (i.e., the ACT) in a high school student population?" . The first part of this chapter includes the CFA used to verify the theoretical model established with a university population in the pilot study. The second part of this chapter includes the Hierarchical Multiple Regressions to predict the relationship between the Independent and Dependent Variables. The following findings are reported: (1) Sample demographic information (2) Outliers and assumptions (3) CFA item descriptive statistics and correlations, (4) CFA main analysis (5) Hierarchical Multiple Regression Outliers and Assumptions and (6) Hierarchical Multiple Regression Results.

Analysis Sample Demographic Information: Main Study

Blanchard and Osborne (2010) suggest that investigators be aware of indiscriminate responses from research participants. Study participants who are not interested in the outcome may provide random answers that are a threat to the validity of research. In order to protect against random responses, the data were screened and resulted in the elimination of invalid responses. A total of 402 high school 11th grade students completed the TEOP. Items in the demographic portion of the measure gender, race, proxies for socioeconomic status, extracurricular participation, hours per day listening to music, do they exercise, and days per week playing video games number of times taking ACT, ACT study guide utilization, cumulative GPA, favorite subject, if they take AP/Honors courses, are they a good test taker, do they have difficulty staying focused when testing, do they study in silence, do they study with food/drink/gum, preferred testing time congruency, “Action” score, “Sound” score, and the dependent variable scores of ACT Composite, ACT Math, ACT Science, ACT English, and ACT Reading. Demographic and other characteristics of the final analysis sample are provided in Table 14.

A total of 51 participants were dropped from the study for inappropriate responses. Forty-two participants were dropped for an inappropriate student identification code. Eight were eliminated for not answering or incompletely answering the TEOP items. Finally, an individual case was dropped for describing their gender as an “Apache Helicopter.” After removal of the 51 cases, a total of 351 participants remained.

Outliers and Assumptions: Main Study CFA

Extreme values (i.e., outliers) are cases that fall outside the normal range on one or more variables. These outliers can cause a distortion (e.g., skewness) in the data set. The presence of an outlier can lead to an increased risk of committing a Type I or Type II Error (Lomax & Hahs-Vaughn, 2012). Outliers are most likely not an issue as the TEOP is scored on a 5- point Likert Scale.

In order to conduct CFA, certain multivariate assumptions must be met. These assumptions include: (1) Normality, (2) Sufficient Sample size, (3) the correct a priori model specification, and (4) Data must be drawn a random sample. Normality was checked prior to conducting the CFA by examining skewness for each item. Six of the nine items were significantly skewed ($p < .001$ for all). The sample size should be large enough to yield reliable estimates of correlations among the variables. Nunnally (1994) suggests that 10 subjects per item is necessary to reduce sampling error. Thus, the sample size of 351 participants was considered acceptable. The model was evidenced to be accurate a priori through the cross-validation process in the Pilot Study. The study sample was drawn from five distinct high schools from four distinct communities. The high schools the participants attended were both public and private. The results for an individual had no impact on anyone else's responses.

Confirmatory Factor Analysis (CFA) Descriptives: Main Study

In the CFA sample ($N = 351$), 169 (48.1%) were male and 182 (51.9%) were female. The sample was comprised of 291 (82.9%) White/Caucasian participants. A majority of the participants (62.4%) attended a private high school. A Bachelor's degree was earned by 113 (32.2%) of the participants' mothers and 97 (27.6%) of their fathers. One hundred and fifty four (43.9%) of the students participated in three or more extra-curricular activities. Two hundred and fifty (71.2%) of the participants exercised. One hundred and eighty four (52.4%) of the respondents believe they are good test takers. A majority or 204 (58.1%) participants feel they have trouble staying focused when testing. Two hundred and fifty four (72.4%) of the respondents preferred taking a test at a time that was not congruent with the ACT's defined start time. The mean GPA was 3.10 (*SD* = .73). The ACT attempt that was taken concurrently with this study was the first attempt for 213 (60.7%) of the respondents. In order to prepare for the ACT exam 214 (61.0%) of the students in the study used a study guide. One hundred and ninety eight (56.4%) of participants have taken an honors or AP class. The favorite subject of 206 (58.7%) of the students is Math, Science, or English; all ACT tested categories, compared to 143 (40.7%) participants who prefer untested subjects Social Studies, Art, or Physical Education. The mean score for the "Action" factor was 3.48 (*SD* = .97) and 2.32 (*SD* = .93) for the "Sound" factor. The mean ACT Composite score was 21.45 (*SD* = 5.41). The mean ACT Math score was 21.18 (*SD* = 5.24). The mean ACT Science score was 21.62 (*SD* = 5.19). The mean ACT English score was 20.93 (*SD* = 6.54). The mean ACT Reading score was 21.95 (*SD* = 6.40).

Table 14

Main Study Demographics (*N =351*)

Variables	*M*(*SD*) or *n*(%)	Regression	*M(SD)* or *n*(%)	Coding
Gender				
Female	182(51.9)	SAME	SAME	0
Male	169(48.1)	SAME	SAME	1
Ethnicity				
White/Caucasian	291(82.9)	Other	60(17.1)	0
Black/African American	29(8.3)	White	291(82.9)	1
American Indian/Alaskan Native	1(.3)			
Asian	5(1.4)			
Hispanic/Latino	11(3.1)			
Bi or Multiracial	14(4.0)			
High School Type				
Public	132(37.6)	SAME	SAME	0
Private	219(62.4)	SAME	SAME	1
Mother's Education				
Not Sure	19(5.4)			
No High School Diploma	17(4.8)			
High School Diploma	58(16.5)			
Some College	62(17.7)	< Bachelors	156(44.4)	0
Bachelor's	113(32.2)	≥ Bachelors	193(55.0)	1
Master's	60(17.1)			
Professional	8(2.3)			
Doctorate	12(3.4)			
Missing	2(.6)			
Father's Education				
Not Sure	32(9.1)			
No High School Diploma	14(4.0)			
High School Diploma	78(22.2)			
Some College	56(16.0)	< Bachelors	180(51.3)	0
Bachelor's	97(27.6)	≥ Bachelors	169(48.1)	1
Master's	44(12.5)			
Professional	14(4.0)			
Doctorate	14(4.0)			
Missing	2(.6)			
Number of ECAs				
0	65(18.5)	No ECAs	65(18.5)	0
1	66(18.8)	= 1 ECAs	66(18.8)	1
2	66(18.8)	= 2 ECAs	66(18.8)	2
3	53(15.0)	= 3 ECAs	53(15.0)	3
4	43(12.3)	= 4 ECAs	43(12.3)	4
5	26(7.4)	≥ 5 ECAs	58 (16.5)	5
6	14(4.0)			
7	9(2.6)			

8	1(0.3)			
9	3(.9)			
10	5(1.4)			
Exercise				
No	100(28.5)	SAME	SAME	0
Yes	250(71.2)	SAME	SAME	1
Missing	1(.3)	--	--	
ACT Attempts				
1	213(60.7)	= 1 Attempt	213(60.7)	0
2	98(27.9)	> 1 Attempt	138 (39.3)	1
3	34(9.7)			
4	5(1.4)			
5	0(0.0)			
6	1(.3)			
GPA (4.0 Scale)	3.10(.73)	--	--	--
Taken an AP/Honors Class				
No	153(43.6)	No	153(43.6)	0
Yes	198(56.4)	Yes	198(56.4)	1
Favorite Subject				
Math	77(21.9)	SS+Art+PE	143(40.7)	0
Science	80(22.8)	MA+SC+Eng	206(58.7)	1
Social Studies	63(17.9)			
English	49(14.0)			
Art	44(12.5)			
Physical Education	36(10.3)			
Missing	2(.6)			
Good Test Taker				
No	165(47.0)	SAME	SAME	0
Yes	184(52.4)	SAME	SAME	1
Missing	2(.6)	--	--	
Trouble Focusing when Testing				
No	146(41.6)	SAME	SAME	0
Yes	204(58.1)	SAME	SAME	1
Missing	1(.3)	--	--	
Test Time Congruency with ACT				
No	254(72.4)	SAME	SAME	0
Yes	97(27.6)	SAME	SAME	1
Action	3.48(.97)	--	--	--
Sound	2.32(.93)	--	--	--
Main Dependent Variables		--	--	--
ACT Composite	21.45(5.41)	--	--	--
ACT Math	21.18(5.24)	--	--	--
ACT Science	21.62(5.19)	--	--	--
ACT English	20.93(6.54)	--	--	--
ACT Read	21.95(6.40)	--	--	--

Note. ECA = Extra-Curricular Activities. SS = Social Studies. PE = Physical Education. MA = Math. SC = Science. Eng = English.

Table 15

The Test Environment for Optimal Performance Items (N = 9)

#	Item
1.	When taking a test, I prefer silence.
2.	When taking a test, I prefer background noise/environmental sounds (e.g., sounds of the ocean, rain, birds chirping).
3.	When taking a test, I prefer to listen to music.
4.	When taking a test, I prefer to have a television on in the background.
5.	When taking a test, I prefer to have water/beverage available.
6.	When taking a test, I prefer to have food available.
7.	When taking a test, I prefer to chew gum or candy.
8.	When taking a test, I prefer to have the option stand up.
9.	When taking a test, I prefer to have the option to walk around.

Item Descriptives and Correlations

In the CFA sample, descriptives for all items are found in Table 16. The item with the highest mean was Item 5 (i.e. "I prefer to have water/a beverage available." $M = 4.10$, $SD = 1.08$), and the item with the lowest mean was Item 4 (i.e., "I prefer to have a television on in the background." $M = 1.77$, $SD = 1.10$). Of the significant correlations, the relationship between Item 8 (i.e., "I prefer to have the option to stand up.") and Item 9 (i.e., "I prefer to have the option to walk around.") was the highest ($r = .735$, $p < .001$). The lowest correlation was between Item 2 (i.e., "I prefer background noise/environmental sounds.") and Item 5 (i.e. "I prefer to have water/a beverage available."; $r = .136$, $p < .05$). The correlations between all the items are presented in Table 17.

Table 16

Main Study TEOP Item Descriptives (N = 351)

Items	*M*	*SD*	*Mdn*	IQR	Skew	Kurt
1. I prefer silence.	3.85	1.13	4.00	2	-.66	-.41
2. I prefer background/noise environmental sounds.	2.46	1.20	3.00	2	.31	-.85
3. I prefer to listen to music.	2.91	1.44	3.00	2	.04	-1.32
4. I prefer to have a television on in the background.	1.77	1.10	1.00	1	1.36	.97
5. I prefer to have water/a beverage available.	4.10	1.08	4.00	2	-1.06	.36
6. I prefer to have food available.	3.54	1.34	4.00	2	-.43	-1.02
7. I prefer to chew gum/candy.	3.83	1.26	4.00	2	-.83	-.33
8. I prefer to have the option to standup.	3.05	1.39	3.00	2	.01	-1.22
9. I prefer to have the option to walk around.	2.87	1.39	3.00	2	.17	-1.18

Note. Min = 1 and Max = 5 for all items.

Table 17

Main Study Inter-Item Spearman Correlation Matrix for the TEOP Items (N = 351)

Item	1	2	3	4	5	6	7	8	9
1. Silence	-	$.491^{***}$	$.528^{***}$	$.410^{***}$	$.117^{*}$	$.226^{***}$	$.140^{**}$	$.209^{***}$	$.215^{***}$
2. Background Noise		-	$.383^{***}$	$.395^{***}$	$.136^{*}$	$.197^{***}$	.067	$.215^{***}$	$.221^{***}$
3. Music			-	$.410^{***}$	$.299^{***}$	$.414^{***}$	$.342^{***}$	$.311^{***}$	$.324^{***}$
4. Television				-	.011	$.217^{***}$	$.137^{*}$	$.165^{**}$	$.217^{***}$
5. Water/Beverages					-	$.596^{***}$	$.466^{***}$	$.334^{***}$	$.341^{***}$
6. Food						-	$.536^{***}$	$.483^{***}$	$.503^{***}$
7. Chew Gum/Candy							-	$.270^{***}$	$.283^{***}$
8. Stand Up								-	$.735^{***}$
9. Walk Around									-

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

The CFA was conducted using LISREL 9.10 (Joreskog & Sorbom, 2013). The results of the CFA were used to evaluate fit of the preliminary factor structure and to assess the underlying factor structure of the instrument. The intention of the factor analytic approaches in the CFA is to generate and validate a factor structure using the nine testing environmental preferences found in the TEOP. Inter-item correlations, the factors, and the factor loadings were examined to evidence the underlying structure of environmental preferences represented by the items in the measure.

CFA Analysis Main Study

In the CFA results are presented in Table 18. For the inter-item correlations, Items 1 through 4 (i.e., the “Sound” factor) were all significantly and positively correlated with each other ($p < .001$ for all). The correlations between items within the “Action” factor (i.e., Items 5-9) were all significantly and positively correlated with each other as well ($p < .001$ for all). The initial CFA showed the observed variables water, food, gum, stand, and walk all loaded significantly ($p < .001$ for all) on the factor “Action.” Item 6 (i.e., “When taking a test, I prefer to have food available.”) had the strongest loading ($\beta = .80$; 63.4% of the variance explained) and Item 7 (i.e., “When taking a test, I prefer to chew gum or candy.”) had the weakest loading ($\beta = .56$; 31.2% of the variance explained).

The observed variables silence, noise, music, and TV all loaded significantly ($p < .001$ for all) on the factor “Sound,” with Item 3 (i.e., “When taking a test, I prefer to listen to music.”) having the strongest loading ($\beta = .73$; 53.5% of the variance explained) and Item 4 (i.e., “When taking a test, I prefer to have a television on in the background.”) with the weakest loading ($\beta = .57$; 32.8% of the variance explained). In the initial model there were nine factor loadings, nine measurement errors, and one factor correlation. Because the distinct values (i.e., unique values) in the matrix *S* (45) are greater than the total number of free parameters (19), this initial model is considered over-identified (i.e., there is more than one way of estimating parameters; (Schumacker & Lomax, 2016).

The χ^2 Goodness-of-Fit statistic was significant (χ^2 [26] = 221.15, $p < .001$). From the other fit indices, the Root-Mean-Square-Error of Approximation (RMSEA) was .146, the Standardized-Root-Mean Residual (SRMR) was .077, the Goodness- of- Fit Index (GFI) was .864, and the Adjusted Goodness-of-Fit (AGFI) Index was .764. Overall, the model did not fit the data. The modification indices were consulted, and a second model was run that allowed the error covariance of Item 8 (i.e., "I prefer to have the option to stand up.") and Item 9 (i.e., "I prefer to have the option to walk around.") to correlate. Additionally, these two items were located within the same "Action" factor, and conceptually, the behaviors of standing and walking are related actions (Jensen, 1998; Schumaker & Lomax, 2016).

The final CFA model showed the observed variables water, food, gum, stand, and walk all loaded significantly ($p < .001$) on the factor "Action," with Item 6 (i.e., "When taking a test, I prefer to have food available.") having the strongest loading (β =.89; 79.9% of the variance explained), and Item 8 (i.e., "When taking a test, I prefer to have the option to stand up.") with the weakest loading (β = .53; 28.3% of the variance explained). The factor "Action" had a direct positive effect on preference for water, food, gum/candy, standing up, and walking around when testing. This predicts that as a participant's "Action" preference increases their inclination toward having water, food, gum/candy, as well as being able to stand up or walk around during testing will also increase. The strong loading of preference to have food available within the "Action" factor suggests that eating is seen as a deed and participants with greater scores on the "Action" items where inclined to prefer accessibility to food. The weaker loading on the preference to stand up indicates that a lower score correlation between this item and the rest of the "Action" items. This may be the result of the error covariance between Item 8 and Item 9 being allowed to correlate.

The observed variables silence, noise, music, and TV all loaded significantly ($p < .001$) on the factor "Sound", with Item 3 (i.e., "When taking a test, I prefer to listen to music.") having the strongest loadings ($\beta = .73$; 53.6% of the variance explained), and Item 4 (i.e., "When taking a test, I prefer to have a television on in the background.") having the weakest loading ($\beta = .57$; 32.9% of the variance explained). The factor "Sound" had a direct positive effect on preference for silence, background noise, music and television when testing. This predicts that as a participant's "Sound" preference increases their penchant for having silence, background noise, music, and television on during testing will increase. The strong loading for a preference of listening to music when testing suggests that those with a high sound preference had higher scores on this item. This could be due to music being strongly identified with sounds and makes conceptual sense. The weaker loading on the preference for having a television on could be due to visual element provided by a television not present in the other items. The positive loading scores on the "Sound" items explain that as the participants' preference for more "Sound" increases their scores on these items increases. In the final model there were nine factor loadings, nine measurement errors, one factor correlation and one error covariance. Because the distinct values (i.e., unique values) in the matrix *S* (45) are greater than the total number of free parameters (20), this final model is considered over-identified.

The χ^2 fit statistic was significant (χ^2 [25] = 73.22, $p < .001$). In addition, the RMSEA was .074, the SRMR was .062, the GFI was .955, and the AGFI was .919. Overall, the majority model fit indices suggested a good fit and there were no additional modifications (Schumaker & Lomax, 2016). The statistically significant is particularly sensitive to sample size. Samples in excess of 200 tend to result in significant χ^2 (Schumacker & Lomax, 2016). Schumaker and Lomax (2016) recommend a majority of the fit indices used designate an acceptable model, than the data supports the theoretical model. The two-factor structure of the TEOP used in the CFA produced evidence of high internal consistency reliability ($\alpha = .813$). The five-item "Action" factor had a higher reliability ($\alpha = .800$) compared to the four-item "Sound" factor ($\alpha = .753$). There were positive correlations between the two subscales (i.e., “Sound” and “Action”) in both the initial model ($r = .495$, $p = .054$) and the final model ($r = .465$, p =.055). This slight correlation indicates a relationship between “Action” and “Sound” scores, though it is not statistically significant.

Table 18

Main Study Standardized Factor Loadings for the TEOP (N = 351)

Item	Initial Model		Model 1	
	F	β	F	β
I prefer to have water/a beverage available.	Action	.62	Action	.67
I prefer to have food available.	Action	.80	Action	.89
I prefer to chew gum/candy.	Action	.56	Action	.61
I prefer to have the option to stand up.	Action	.69	Action	.53
I prefer to have the option to walk around.	Action	.71	Action	.55
I prefer silence.	Sound	.72	Sound	.73
I prefer background noise/environmental sounds.	Sound	.61	Sound	.60
I prefer to listen to music.	Sound	.73	Sound	.73
I prefer to have a television on in the background.	Sound	.57	Sound	.57

Note. F = Factor for each item (i.e., Action and Sound); β = the standardized factor loadings; Initial Model = the original 2-factor; Model 1 = Error covariance was added between Item 8 and Item 9; CFA= Confirmatory Factor Analysis.

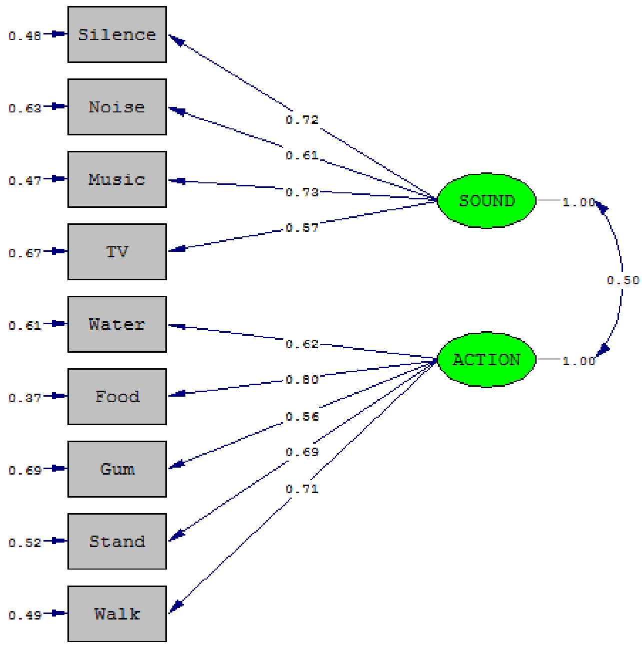

Figure 4. Initial model for the Confirmatory Factor Analysis (CFA) Main Study.

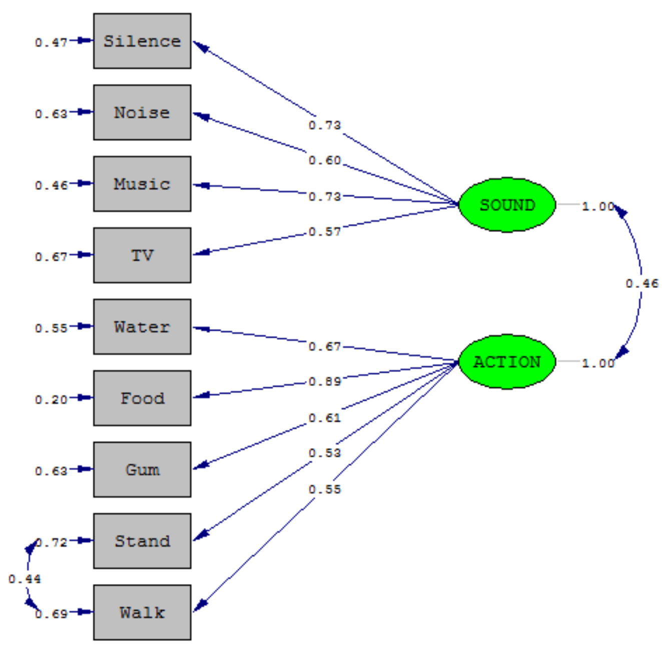

Figure 5. Final model for the Confirmatory Factor Analysis (CFA) Main Study.

CFA Discussion (Main Study)

The objective of this study was to examine current high school students', specifically high school juniors, perceptions of preferred standardized testing conditions and what is the relationship with standardized test scores, specifically the ACT. The primary research questions that guided the study were, (1) "What are the psychometric properties (i.e., construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a high school student population?," and (2) "What is the relationship between the TEOP factor scores (i.e., Action and Sound) and high-stakes aptitude test scores (i.e., the ACT) in a high school student population?"

Confirmatory Factor Analysis (CFA)

A CFA was conducted to test if the data fit the hypothesized factor structure. After consulting the modification indices, the model included the observed variables of water, food, gum, stand, and walk all loading significantly on the "Action" factor. Item 6 (i.e., "I prefer to have food available.") had the strongest loading and the "Action" factor explained the largest proportion of variance in this item. This can be attributed to the deliberate steps necessary to bring food to the testing location. Additionally, Item 8 (i.e., "I prefer to have the option to stand up.") had the weakest loading and the "Action" factor explained the least amount of variance in this item. Item 8's loading and amount of explained variance both decreased considerably from the initial model to the final model. This reduction can be attributed to the added error covariance between Items 8 and 9.

The observed variables of silence, noise, music, and TV all loaded significantly on the factor “Sound.” Item 3 (i.e., “I prefer to listen to music.”) had the strongest loadings and the “Sound” factor explained the largest proportion of variance in this Item. The strong loading can be explained by the direct association between music and sound. Item 4 (i.e., “I prefer to have television on in the background.”) had the weakest loading and the “Sound” factor explained the smallest proportion of variance in this item. The low loading could be attributed to television being perceived as a visual medium more so than sound.

The largest mean score from the TEOP “Action” items in the CFA sample was Item 6 (i.e., “I prefer to have food available.”) and the lowest score was on Item 8 (i.e., “I prefer to have the option to stand up.”). The high preference for being able to eat indicates that test takers would like the option of having food during testing to perhaps the deliberate movements of eating such as bring the food to the mouth, chewing, and swallowing, as well as, mitigating effect on hunger resonate with test takers with “Action” preferences. The lower preference for standing up may be due to the timed nature of these exams. Students may be less inclined to consider standing up during a test, as this would require subtracting time from their allotted examination period. Students may be anxious or concerned that standing up will prevent them from finishing the test. Finally, the highest mean score for “Sound” was Item 1 (i.e., “I prefer silence.”) and the lowest was Item 4 (i.e., “I prefer to have a television on in the background.”). As with the Pilot sample, most participants prefer silence, and few favor noisy appliances/devices with visual distractions.

The inter-item correlations within the "Action" factor (i.e., Items 5 to 9) were all significant and positive in direction. The same was noted for Items 1 through 4 in the "Sound" factor. Additionally, stronger relationships were noted among items within the same factor compared to items located in the other factor. The strongest correlation among the "Action" items was between Item 8 (i.e., "I prefer to have the option to stand up.") and Item 9 (i.e., "I prefer to have the option to walk around."). Similarly for the "Sound" factor, Item 1 (i.e., "I prefer silence.") and Item 3 (i.e., "I prefer to listen to music.") had the strongest correlation. The two-factor TEOP structure was confirmed, with the five-item "Action" factor having higher internal consistency reliability compared to the four-item "Sound" factor. The CFA internal consistency reliabilities for the two factors were opposite in order of magnitude compared to the EFA reliabilities. For the CFA, the factor with more items had higher internal consistency reliability. Overall, the internal consistency reliability increased for both factors in the Main Study CFA sample compared to the Pilot Study.

Summary of Confirmatory Factor Analysis

Further examination of the item correlation matrices revealed a near "textbook" example of the within and between patterns of association that should be rendered after following best practices in construct-based measure construction. For "Action," all the items contained within that factor had significant, positive, and high correlations with each other. These items were also not as strongly correlated with the items in the "Sound" factor. The correlations between items located in different factors (e.g., a correlation coefficient between an item in the "Action" factor and an item in the "Sound" factor) were inconsistent compared to the correlations amongst the items contained within each factor (e.g., a correlation coefficient between two items within the “Action” factor or between two items within the “Sound” factor). That is, there was less variation in the within-factor item correlations compared to the variation in the between-factor item correlations. The correlation patterns followed the expected template of between and within group relationships.

Hierarchical Multiple Linear Regression Results

Hierarchical Multiple Linear Regressions were conducted to analyze Research Question 2 (RQ2), which states: “What is the relationship between the TEOP factor scores (i.e., Action and Sound) and high-stakes aptitude test scores (i.e., the ACT) in a high school student population?” Five regressions were run, using different dependent variables for each (i.e., ACT Comprehensive, ACT Math, ACT Science, ACT English, and ACT Reading).

Data Cleaning

In addition to the eliminated cases mentioned previously, cases were also excluded in the second research question’s final analysis sample. There were 47 additional cases removed from RQ2 with missing data on any IV. After the removal of these cases, the final analysis sample for RQ2 was 304.

Demographics

In the analysis sample ($N = 304$), 146 (48.0%) were male and 158 (52.0%) were female. The sample was comprised of 258 (84.9%) White/Caucasian participants. A slight majority of the participants ($n = 195$, 64.1%) attended a private high school. A

Bachelor's degree or higher was earned by 182 (59.9%) of the participants' mothers and 164 (53.9%) of their fathers. The mean number of extra-curricular activities the sample participated in was 2.32 (*SD* = 1.73). Two-hundred twenty three (73.4%) of the respondents stated that they exercised.

One-hundred sixty seven (54.9%) of the respondents believed they are "good test takers." A majority of participants (*n* =170; 55.9%) felt that they have trouble staying focused when testing. Two-hundred eighteen (71.7%) of the respondents preferred taking a test at a time that was not congruent with the ACT's defined morning start time. The mean GPA for the sample was 3.18 (*SD* = .70). The ACT test that was taken concurrently during this study was the first attempt for 172 (56.6%) of the respondents. One-hundred eighty one (59.5%) of participants took at least one AP or honors class.

The favorite subjects of 183 (60.2%) of the students were Math, Science, or English. These subjects were all ACT tested areas, compared to 121 (39.8%) participants who preferred untested subjects such as Social Studies, Art, or Physical Education. Finally, for the main IVs, the mean for the "Action" factor was 3.46 (*SD* = .97) and 2.30 (*SD* = .92) for the "Sound" factor. The mean ACT Composite score was 21.95 (*SD* = 5.30). For the subject-specific ACT tests, the mean ACT Math, Science, English, and Reading scores were 21.63 (*SD* = 5.23), 22.00 (*SD* = 5.13), 21.46 (*SD* = 6.57), and 21.41 (*SD* = 6.41), respectively.

Table 19

Research Question 2 Demographic Information (N = 304)

Variables	*M*(*SD*) or *n*(%)	Coding
Gender		
Female	158(52.0)	0
Male	146(48.0)	1
Ethnicity		
Other	46(15.1)	0
White	258(84.9)	1
High School Type		
Public	109(35.9)	0
Private	195(64.1)	1
Mother's Education		
< Bachelors	122 (40.1)	0
≥ Bachelors	182(59.9)	1
Father's Education		
< Bachelors	140 (46.1)	0
≥ Bachelors	164(53.9)	1
Number of ECAs	2.32(1.73)	--
Exercise		
No	81(26.6)	0
Yes	223(73.4)	1
Good Test Taker		
No	137 (45.1)	0
Yes	167 (54.9)	1
Trouble Focusing on Test		
No	134 (44.1)	0
Yes	170 (55.9)	1
Test Time Congruent with ACT		
No	218 (71.7)	0
Yes	86 (28.3)	1
GPA (4.0 Scale)	3.18 (.70)	--
ACT Attempts		
=1	172 (56.6)	0
≥2	132 (43.4)	1
AP Honors		
No	123 (40.5)	0
Yes	181(59.5)	1
Favorite Class		
SS+Art+PE	121 (39.8)	0
MA+SC+Eng	183 (60.2)	1
Action	3.46 (.97)	--
Sound	2.30 (.92)	--
ACT Comp	21.95 (5.30)	--
ACT Math	21.63 (5.23)	--

ACT Science	22.00 (5.13)	--
ACT English	21.46 (6.57)	--
ACT Reading	21.41 (6.41)	--

Note. ECA = Extra-Curricular Activities. SS = Social Studies. PE = Physical Education. MA = Math. SC = Science. Eng = English.

Outliers and Assumptions – ACT Comprehensive

For the Hierarchical Multiple Regression with ACT Comprehensive Score as the Dependent Variable (DV), centered leverage, Cook's Distance (*D*), and Mahalanobis Distance values were examined to identify any potential outliers or influential cases in the model. The centered leverage values, which are a measure of distance from one case to the average of the independent variables, did not suggest any problematic data (Lomax & Hahs-Vaughn, 2012). The minimum centered leverage value was .018 and the maximum was .097 with the average at .053. Additionally, the centered leverage values were less than .2, indicating no extreme scores for any of the variables. Cook's *D*, an overall measure of cases' individual influence, indicated that all values were close to zero (i.e., values closer to 1 or 2 indicate potential undue influence on the model). The minimum was zero and the maximum was .033, with the average value at .003.

Mahalanobis Distance was consulted, which is a measure of the distance from each case to the mean of the independent variable for the remaining cases (i.e., multivariate outlier detection; Lomax & Hahs-Vaughn, 2012). Typically, this diagnostic is evaluated conservatively ($p < .001$) using the Chi-Square distribution with the degrees of freedom equivalent to the number of predictors in the model ($df = 15$). With an alpha level of .001, the critical value was 30.578. Any distance greater than this value suggests that a case is an outlier. The maximum value was 29.525, with an average of 15.947.

Thus, as no values were greater than 30.578, there was no evidence to suggest any multivariate outliers in the model.

Before examining the ACT Comprehensive regression model, the basic assumptions of Multiple Linear Regression were investigated. To examine the assumption of independence, scatterplots of studentized residuals against the predicted values and IVs were investigated. For each scatterplot, a random display of points between -2 and +2 was observed.

For Linearity, separate partial regression plots were examined for each IV and the DV. A random display of data points was observed falling approximately between the boundaries of -2 and +2. These residual scatterplots of Y and $\hat{Y}$ indicated that linearity was met. Homoscedasticity was examined with scatterplots of the IVs and the DV. The dispersion of the values around the regression line remained fairly constant for all values of X. The assumption of Normality was examined using skewness and kurtosis statistics and histograms of the residuals. The errors appeared to be normally distributed, thus the assumption was met. Finally, Multicollinearity was examined via a correlation matrix. There were no unusually high correlation coefficients ($r \leq .80$). The tolerances for all the predictors were within acceptable limits with the Variance Inflation Factors (VIFs) corroborating this evidence. The collinearity diagnostics did not indicate any overlap in the contribution of the percentages of variance explained to the model.

Outliers and Assumptions - ACT Math

For the Hierarchical Multiple Regression with ACT Math as the DV, centered leverage, Cook's Distance (*D*), and Mahalanobis Distance values were examined. The centered leverage values did not indicate any problematic data (Lomax & Hahs-Vaughn, 2012). The minimum centered leverage value was .018 and the maximum was .097 with the average at .053. Additionally, the centered leverage values were less than .2, indicating no extreme scores for any of the variables. Cook's *D* indicated that all values were close to zero, with the minimum at zero and the maximum at .032 (i.e., the average was .003). Mahalanobis Distance was consulted, and with an alpha level of .001, the critical value was 30.578. The maximum value was 29.525, with an average of 15.947. Thus, as no values were greater than 30.578, there was no evidence to suggest that there were any multivariate outliers in the model.

Before examining the ACT Math regression model, the basic assumptions of Multiple Linear Regression were investigated. For independence, the scatterplots of studentized residuals depicted a random display of points between -2 and +2. For Linearity, separate partial regression plots showed a random display of data points falling approximately between the boundaries of -2 and +2, indicating that the assumption was met. Homoscedasticity was examined with scatterplots, and the dispersion of the values around the regression line remained fairly constant for all values of X. For the assumption of Normality, the histograms of residuals appeared to be normally distributed indicating that the assumption was met. Finally, for Multicollinearity, the correlation matrix showed no unusually high coefficients ($r \leq .80$). The tolerances for all the predictors were within acceptable limits with the VIFs corroborating this evidence. The collinearity diagnostics did not indicate any overlap in the contribution of the percentages of variance explained to the model.

Outliers and Assumptions - ACT Science

For the Hierarchical Multiple Regression with ACT Science as the DV, centered leverage, Cook's Distance (*D*), and Mahalanobis Distance values were examined. The centered leverage values did not suggest any problematic data (Lomax & Hahs-Vaughn, 2012). The minimum centered leverage value was .018 and the maximum was .097 with the average at .053. Additionally, the centered leverage values were less than .2, indicating no extreme scores for any of the variables. Cook's *D* indicated that all values were close to zero with the minimum was zero and the maximum at .040,(i.e., the average was .003).

Mahalanobis Distance was consulted, and with an alpha level of .001, the critical value was 30.578. The maximum value was 29.525, with an average of 15.947. Thus, as no values were greater than 30.578, there was no evidence to suggest that there were any multivariate outliers in the model.

Before examining the ACT Science regression model, the basic assumptions of Multiple Linear Regression were investigated. For independence, the scatterplots of studentized residuals depicted a random display of points between -2 and +2..

For Linearity, separate partial regression plots showed a random display of data points falling approximately between the boundaries of -2 and +2, indicating that the assumption was met. Homoscedasticity was examined with scatterplots, and the dispersion of the values around the regression line remained fairly constant for all values of X. For the assumption of Normality, the histograms of residuals appeared to be normally distributed indicating that, the assumption was met. Finally, for

Multicollinearity the correlation matrix showed no unusually high coefficients ($r \leq .80$). The tolerances for all the predictors were within acceptable limits with the VIFs corroborating this evidence. The collinearity diagnostics did not indicate any overlap in the contribution of the percentages of variance explained to the model.

Outliers and Assumptions - ACT English

For the Hierarchical Multiple Regression with ACT English as the DV, centered leverage, Cook's Distance (*D*), and Mahalanobis Distance values were examined. The centered leverage values did not suggest any problematic data (Lomax & Hahs-Vaughn, 2012). The minimum centered leverage value was .018 and the maximum was .097 with the average at .053. Additionally, the centered leverage values were less than .2, indicating no extreme scores for any of the variables. Cook's *D* indicated that all values were close to zero with the minimum was zero and the maximum at .032, (i.e., the average was .003).

Mahalanobis Distance was consulted, and with an alpha level of .001, the critical value was 30.578. The maximum value was 29.525, with an average of 15.947. Thus, as no values were greater than 30.578, there was no evidence to suggest that there were any multivariate outliers in the model.

Before examining the ACT English regression model, the basic assumptions of Multiple Linear Regression were investigated. For independence, the scatterplots of studentized residuals depicted a random display of points between -2 and +2.

For Linearity, separate partial regression plots showed a random display of data points falling approximately between the boundaries of -2 and +2, indicating that the assumption was met. Homoscedasticity was examined with scatterplots, and the dispersion of the values around the regression line remained fairly constant for all values of X. For the assumption of Normality, the histograms of residuals appeared to be normally distributed indicating that, the assumption was met. Finally, for Multicollinearity the correlation matrix showed no unusually high coefficients ($r \leq .80$). The tolerances for all the predictors were within acceptable limits with the VIFs corroborating this evidence. The collinearity diagnostics did not indicate any overlap in the contribution of the percentages of variance explained to the model.

Outliers and Assumptions - ACT Reading

For the Hierarchical Multiple Regression with ACT Reading as the DV, centered leverage, Cook's Distance (*D*), and Mahalanobis Distance values were examined. The centered leverage values did not suggest any problematic data (Lomax & Hahs-Vaughn, 2012). The minimum centered leverage value was .018 and the maximum was .097 with the average at .053. Additionally, the centered leverage values were less than .2, indicating no extreme scores for any of the variables. Cook's *D* indicated that all values were close to zero with the minimum was zero and the maximum at .037(i.e., the average was .003).

Mahalanobis Distance was consulted, and with an alpha level of .001, the critical value was 30.578. The maximum value was 29.525, with an average of 15.947. Thus, as no values were greater than 30.578, there was no evidence to suggest that there were any multivariate outliers in the model.

Before examining the ACT Reading regression model, the basic assumptions of Multiple Linear Regression were investigated. For independence, the scatterplots of studentized residuals depicted a random display of points between -2 and +2.

For Linearity, separate partial regression plots showed a random display of data points falling approximately between the boundaries of -2 and +2, indicating that the assumption was met. Homoscedasticity was examined with scatterplots, and the dispersion of the values around the regression line remained fairly constant for all values of X. For the assumption of Normality, the histograms of residuals appeared to be normally distributed indicating that, the assumption was met. Finally, for Multicollinearity the correlation matrix showed no unusually high coefficients ($r \leq .80$). The tolerances for all the predictors were within acceptable limits with the VIFs corroborating this evidence. The collinearity diagnostics did not indicate any overlap in the contribution of the percentages of variance explained to the model.

Correlations were run to examine the relationship between IVs and DVs. Nonparametric Spearman's Rho Correlations were computed for the relationships between ordinal and ordinal variables. Point Biserial Correlations were run between dichotomous and interval-level variables. Phi Correlations were calculated between dichotomous variables, and Pearson Correlations were run between interval-level variables. The highest positive correlation among the IVs was between GPA and if the student took an AP or honors course ($r_{pb} = .603$, $p < .001$), and the highest negative correlation was between students believing they are a "good test taker" and their ability to focus when testing ($r_{\varphi} = -.464$ $p < .001$). The highest positive correlation indicates that those who take AP and honors courses are more likely to have higher GPAs. The highest negative correlation suggests that students who have more trouble focusing during testing also believe they are “bad test takers.”

Of the IVs, the strongest positive relationship with the DV (ACT Comprehensive) was GPA ($r = .742$, $p < .001$), and the strongest negative relationship with that DV was the ability to focus when testing ($r_{pb} = -.333$, $p < .001$). The highest positive correlation indicates that those with higher GPAs receive higher ACT Comprehensive scores. The highest negative correlation specifies that students who cannot focus when testing get lower ACT Comprehensive scores. Of the IVs, the strongest positive relationship with the DV ACT Math was GPA ($r = .708$, $p < .001$), and the strongest negative relationship was the ability to focus when testing ($r_{pb} = -.282$, $p < .001$). The highest positive correlation suggests that those with good grades receive higher Math scores on the ACT. The highest negative correlation suggests that students who do not focus well when testing get lower Math scores on the ACT.

Of the IVs, the strongest positive relationship with the DV ACT Science was GPA ($r = .662$, $p < .001$), and the strongest negative relationship was the ability to focus when testing ($r_{pb} = -.297$, $p < .001$). The highest positive correlation indicates that those with good grades receive higher Science scores on the ACT. The highest negative correlation suggests that students who do not focus well when testing get lower Science scores on the ACT. Of the IVs, the strongest positive relationship with the DV ACT English was GPA ($r = .738$, $p < .001$), and the strongest negative relationship was the ability to focus when testing ($r_{pb} = -.243$, $p < .001$). The highest positive correlation indicates that those with good grades receive higher English scores on the ACT. The highest negative correlation advocates that students who do not focus well when testing get lower English scores on the ACT.

Of the IVs, the strongest positive relationship with the DV ACT Reading was GPA (r= .631, p < .001), and the strongest negative relationship was the ability to focus when testing (r_{pb} = -.311, p < .001). The highest positive correlation designates that those with good grades receive higher Math scores on the ACT. The highest negative correlation suggests that students who do not focus well when testing get lower Reading scores on the ACT.

GPA had the highest positive correlation with all five DVs, which provides evidence that higher cumulative grades have a strong relationship to students' high-stakes standardized test scores on the ACT. Trouble staying focused when testing had the highest negative correlation with all five DVs. This suggests that test takers who do not focus well during testing may not have higher scores on the ACT compared to those who can concentrate for the duration of a test. The TEOP variable "Action" had a statistically significant and negative correlation (p < .05 for all) with four of the DVs (i.e., ACT Comprehensive, Math, Science, and English), and the correlations ranged in magnitude from -.074 to -.115. This indicates that as participants' preference for "Action" increases (i.e., the preference to physically move around during a test), their scores on the ACT Comprehensive and individual subject matter tests for three of the four categories (i.e., Math, Science, and English) decreases. The TEOP Variable "Sound" had a statistically significant and negative correlation (p < .01 for all) with all five DVs (i.e., ACT

Comprehensive, Math, Science, English, and Reading) ranging from -.139 to -.333. This suggests that as a participant's preference for "Sound" escalates (i.e., the preference to listen to music or other noises), their scores on the ACT Comprehensive and all four subject matter categories decreases.

Table 20

Main Study Regression Correlations (N = 304)

Items	1	2	3	4	5	6	7	8	9	10
1. Gender	--	.119*	.030	.054	.101	-.143**	.089	.079	-.057	-.009
2. Ethnicity		--	-.165**	.111*	.090	.067	.082	.247***	-.123*	.078
3. HS Type			--	.157**	.208***	.204***	.199***	-.082***	.023	.112*
4. Mom Ed				--	.433***	.259***	.175**	.080	-.148**	.016
5. Dad Ed					--	.258***	.124*	.125*	-.143*	.023
6. ECA						--	.248 ***	.106*	-.094*	.186
7. Exercise							--	.068	-.036	.081
8. GTT								--	-.464***	.062
9. Focus									--	-.098
10. Time										--
11. GPA										
12. ACT attempt										
13. AP/Honors										
14. Fav Class										
15. Action										
16. Sound										
17. ACT Comp										
18. ACT Math										
19. ACT Sci										
20. ACT Eng										
21. ACT Read										

Note. *p<.05, **p<.01, ***p<.001; HS = High School; Ed = Education; ECA = Extra-Curricular Activities; GTT = Good Test Taker; GPA = Grade Point Average (4.0 Scale); Focus = Focus when Testing; Time = Test Time Congruency; Comp = Comprehensive; Sci = Science; Eng = English; Read = Reading.

Table 20 (continued)

Main Study Regression Correlations (N = 304)

Items	11	12	13	14	15	16	17	18	19	20	21
1. Gender	-.124*	-.110*	-.142**	-.188***	-.086	.061	-.020	.103	.020	-.091	-.099
2. Ethnicity	.244***	.180**	.257***	.068	-.052	-.045	.275***	.279***	.267***	.265***	.220***
3. HS Type	.184***	.095	-.101	.057	-.149**	-.211***	.081	.023	.045	.128*	.061
4. Mom Ed	.330***	.228***	.219***	.090	-.104	-.142**	.286***	.247***	.268***	.281***	.260***
5. Dad Ed	.278***	.243***	.210***	.033	-.049	-.165**	.345***	.295***	.274***	.368***	.325***
6. ECA	.513***	.366***	.385***	.176**	-.036	-.195***	.402***	.363***	.333***	.414***	.363***
7. Exercise	.183**	.183**	.157**	-.022	-.014	-.110*	.124*	.148**	.107*	.103	.109*
8. GTT	.311***	.167**	.273***	.096	-.071	-.056	.462***	.413***	.414***	.438***	.421***
9. Focus	-.286***	-.171**	-.192***	-.168**	.175**	.178**	-.333***	-.282***	-.297***	-.343***	-.311***
10. Time	.147***	.076	.055	.127*	-.101	-.233***	.088	.092	.063	.126*	.061
11. GPA	--	.490***	.603***	.284***	-.188***	-.333***	.742***	.708***	.662***	.738***	.631***
12. ACT attempt		--	.343***	.173***	-.032	-.225***	.419***	.384***	.368***	.474***	.342***
13. AP/Honors			--	.154**	-.043	-.138**	.620***	.607***	.552***	.574***	.543***
14. Fav Class				--	-.114*	-.118*	.265***	.254***	.215***	.250***	.223***
15. Action					--	.394***	-.114*	-.109*	-.107*	-.115*	-.074
16. Sound						--	-.191***	-.204***	-.160**	-.232***	-.139***
17. ACT Comp							--	.887***	.920***	.928***	.905***
18. ACT Math								--	.815***	.770***	.697***
19. ACT Sci									--	.803***	.784***
20. ACT Eng										--	.819***
21. ACT Read											--

Note. $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$; HS = High School; Ed = Education; ECA = Extra-Curricular Activities; GTT = Good Test Taker; GPA = Grade Point Average (4.0 Scale); Focus = Focus when Testing; Time = Test Time Congruency; Comp = Comprehensive; Sci = Science; Eng = English; Read = Reading.

Hierarchical Multiple Linear Regressions

Five Hierarchical Multiple Linear Regression analyses were conducted to determine if a relationship exists between TEOP factor scores (i.e., Action and Sound) and the ACT Comprehensive test as well as each individual ACT subject matter test (i.e., Math, Science, English, and Reading) in a sample of high school students controlling for demographic/behavioral, study/testing, and high school academic variables.

ACT Comprehensive

A Hierarchical Multiple Linear Regression was conducted that examined the relationship between the TEOP and ACT Comprehensive, controlling for demographic/behavioral, study/testing, and high school academic variables. The results suggest that a significant proportion of the total variation in the DV ACT Comprehensive was predicted by the collection of IVs (R^2 = .636; F[16, 287] = 31.359, p <.001). Together, the predictors accounted for 63.6% of the variance in Comprehensive ACT scores. However, the fourth block of predictors (i.e., TEOP scores), once added to the model, did not produce a significant change in the F statistic (ΔR^2 = .000, p =.896).

The model contained 16 predictors with the TEOP variables as the main IVs (i.e., Action and Sound). Of the 16 IVs in the model, from the first block, two of the seven demographic/behavioral variables were significant predictors of ACT Comprehensive score. These variables were Gender and Father's Level of Education. In the second block, Study/Testing Variables, one of the three IVs was a significant predictor of the outcome variable. This variable was if the student feels that he/she is a good test taker. In the third block, all four of the high school academic variables (i.e., GPA, number of ACT attempts, AP/Honors classes, and if their favorite class was a tested ACT subject) were significant predictors of the DV. In the fourth block, neither of the TEOP variables were significant predictors of ACT Comprehensive scores.

The following paragraph presents the statistical tests for each significant predictor in the model. From the first block, Gender was a statistically significant predictor of ACT Comprehensive (t = 2.237, df = 287, p = .026). Specifically, Gender was positively predictive of ACT Comprehensive scores (β = .088; B = .927, SE = .414). If a student is female, their ACT Comprehensive score will be approximately .927 points higher than a male student. Father's level of education was a statistically significant predictor of ACT Comprehensive (t = 2.859, df = 287, p =.005), and was positively predictive of the outcome (β = .119; B = 1.258, SE = .440). If a student's father had at least a Bachelor's degree, their ACT Comprehensive score was predicted to be approximately 1.258 points higher than those with fathers who have lower levels of education.

From the second block, the variable "Good Test Taker" (t = 5.438, df = 287, p < .001) was a statistically significant predictor of ACT Comprehensive score, and was positively related (β = .233; B = 2.472, SE = .455) to the outcome. For this variable, if a student feels that he/she is a good test taker, his/her predicted ACT Comprehensive score will be 2.472 points higher than those who do not feel that they are good test takers.

From the third block, GPA (t = 7.652, df = 287, p < .001) was a statistically significant predictor of ACT Comprehensive score. Specifically, GPA was positively predictive (β = .417; B = 3.318, SE = .410) of the outcome. For every additional GPA point, students' predicted ACT Comprehensive scores will increase by 3.318 points. The number of ACT attempts (t = 1.127, df = 287, p = .011) was a statistically significant predictor and was positively predictive of ACT Comprehensive scores (β = .106; B = 1.127, SE = .441). Students who took the ACT more than once scored 1.127 points higher than those who were taking the test for the first time. The variable representing AP or Honors course enrollment (t = 2.602, df = 287, p < .001) was a statistically significant and positive predictor of ACT Comprehensive scores. If students took an AP or Honors course (β = .241; B = 2.602, SE = .508), he/she would have a 2.602 point higher score on the ACT Comprehensive (i.e., compared to those who have not taken an AP or Honors course).

Finally, the variable representing if a students' favorite school subject was one of the ACT tested areas (t = 1.1219, df = 287, p = .004) was a statistically significant predictor of ACT Comprehensive scores. If a student's favorite course was on the ACT (β = .113; B = 1.219, SE = .419), his/her Comprehensive test scores were predicted to be 1.219 points higher than those whose favorite course was not on the ACT. For the main variables of interest in the model, both of the TEOP variables "Action" (t = -.076, df = 287, p = .940; β = -.003; B = -.017, SE = .223) and "Sound" (t = .455, df = 287, p = .649; β = .020; B = .115, SE = .253), were not significantly predictive of ACT Comprehensive. A summary of these results can be found in Table 21.

Table 21

Main Study Summary of Hierarchical Multiple Linear Regression Analyses ACT Comprehensive (N = 304)

Independent Variables	*B*	*SE*	*t*	*β*	95% CI for *B*	
					Lower	Upper
Demographic/Behavioral						
Gender	.927	.414	2.237*	.088	.111	1.742
Ethnicity	-.121	.572	-.212	-.008	-1.247	1.005
High School Type	.448	.449	.999	.041	-.435	1.331
Mother's Education	.299	.446	.670	.028	-.579	1.177
Father's Education	1.258	.440	2.859**	.119	.392	2.124
# of Extra-Curricular	-.102	.138	-.740	-.033	-.373	.169
Exercise	-.737	.455	-1.619	-.062	-1.633	.159
Study/Testing						
Good Test Taker	2.472	.455	5.544***	.233	1.577	3.367
Focus when Testing	-.075	.451	-.165	-.007	-.962	.813
Preferred Time Congruency	-.177	.450	-.394	-.015	-1.063	.709
High School Academic						
GPA	3.138	.410	7.652***	.412	2.331	3.945
ACT Attempts	1.127	.441	2.553*	.106	.258	1.995
AP/Honors	2.602	.508	5.111***	.241	1.601	3.602
Favorite Class	1.219	.419	2.910**	.113	.394	2.043
TEOP						
Action	-.017	.223	-.076	-.003	-.456	.423
Sound	.115	.253	.455	.020	-.383	.613

Note. *$p < .05$, **$p < .01$, ***$p < .001$. SS= Social Studies, PE = Physical Education, MA = Math, SC = Science, Eng = English. *B* = unstandardized regression coefficient, *SE* = standard error, β = standardized regression coefficients, CI = confidence interval

ACT Math

A Hierarchical Multiple Linear Regression was conducted that examined the relationship between the TEOP and ACT Math, controlling for demographic/behavioral, study/testing, and high school academic variables. The results suggest that a significant proportion of the total variation in the DV ACT Math was predicted by the collection of IVs ($R^2 = .579$; $F[16, 287] = 24.703$, $p < .001$). Together, the predictors accounted for 57.9% of the variance in the ACT Math scores. However, the fourth block of predictors (i.e., TEOP Scores) once added to the model did not produce a significant change in the *F* statistic (ΔR^2 = .000; *p* = .892).

The model contained 16 predictors with the TEOP variables as the main IVs (i.e., Action and Sound). Of the 16 IVs in the model, from the first block, one demographic/behavioral variable, Gender was a significant predictor of one's ACT Math score. In the second block, Study/Testing Variables one of the three IVs was a significant predictor of the outcome variable. This was variable student feels he/she is a good test taker. In the third block three of the four of the high school academic data variables GPA, if they take AP/Honors classes, and if their favorite class was a tested ACT subject were significant predictors of the DV. In the fourth block neither of the TEOP variables were significant predictors of ACT Math scores.

The following paragraph presents the statistical tests for each significant predictor in the model. From the first block, Gender was a statistically significant predictor of ACT Math (t = 5.825, df = 287, p < .001). Specifically, Gender was positively predictive of ACT Math scores (β = .245; B = 2.560, SE = .439). If a student is male, their ACT Math score will be predicted to be approximately 2.560 points higher than those who are female.

From the second block, the variable "Good Test Taker" (t = 3.513, df = 287, p = .001) was a statistically significant predictor of ACT Math score and was positively related (β = .162; B = 3.513, SE = .482) to the outcome. For "Good Test Taker", if a student feels he/she is a good test taker, his/her predicted ACT Math score will be 1.694 points higher than those who do not feel that they are good test takers.

From the third block, GPA (t = 7.360, *df* = 287, *p* < .001) was a statistically significant predictor of ACT Math score. Specifically, GPA was positively predictive (β = .426; *B* = 3.3202, *SE* = .435) of the outcome results. For every additional GPA point, their predicted ACT Math scores will increase by 3.202 points. The variable representing AP or Honors course enrollment (t = 5.398, *df* = 287, *p* < .001) was a statistically significant and positive predictor of ACT Math scores. If students took an AP or Honors course (β = .274; *B* = 2.911, *SE* = .539) he/she would have 2.911 point better score on the ACT Math (i.e., compared to those who have not taken an AP or Honors course). Finally, the variable representing if a students' favorite school subject was one of the ACT tested areas (t = 3.681, *df* = 287, *p* < .001) was a statistically significant predictor of ACT Math scores. If students' favorite course was on the ACT (β = .153; *B* = 1.635, *SE* = .444) his/her ACT Math scores was predicted to be 1.635 points higher than those whose favorite course was not on the ACT. For the main variables of interest in the model, both of the TEOP variables "Action" (*t* = .471, *df* = 287, *p* = .638) (β = .021; *B* = .112, *SE* = .237), and "Sound" (*t* = -.101, *df* = 287, *p* = .920) (β = -.005; *B* = -.027, *SE* = .268) were not significantly predictive of ACT Math. A summary of these results can be found in Table 22.

Table 22

Main Study Summary of Hierarchical Multiple Linear Regression Analyses ACT Math (N = 304)

Independent Variables	*B*	*SE*	*t*	*β*	95% CI for *B*	
					Lower	Upper
Demographic/Behavioral						
Gender	2.560	.439	5.855***	.245	1.695	3.425
Ethnicity	-.525	.607	-.866	-.036	-1.720	.699
High School Type	-.366	.476	-.770	-.034	-1.303	.571
Mother's Education	.176	.473	.371	.017	-.755	1.107
Father's Education	.545	.467	1.167	.052	-.374	1.463
# of Extra-Curricular	-.107	.146	-.731	-.035	-.395	.181
Exercise	-.169	.483	-.350	-.014	-1.120	.781
Study/Testing						
Good Test Taker	1.694	.482	3.513**	.162	.745	2.643
Focus when Testing	.165	.478	.344	.016	-.777	1.106
Preferred Time Congruency	-.072	.478	-.150	-.006	-1.012	.868
High School Academic						
GPA	3.202	.435	7.360***	.426	2.345	4.058
ACT Attempts	.827	.468	1.767*	.079	-.094	1.749
AP/Honors	2.911	.539	5.398***	.274	1.849	3.972
Favorite Class	1.635	.444	3.681***	.153	.761	2.501
TEOP						
Action	.112	.237	.471	.021	-.355	.578
Sound	-.027	.268	-.101	-.005	-.555	.501

Note. *p < .05, **p < .01, ***p < .001. SS= Social Studies, PE = Physical Education, MA = Math, SC = Science, Eng = English. *B* = unstandardized regression coefficient, *SE* = standard error, β = standardized regression coefficients, CI = confidence interval

ACT Science

A Hierarchical Multiple Linear Regression was conducted that examined the relationship between the TEOP and ACT Science, controlling for demographic/behavioral, study/testing, and high school academic variables. The results suggest that a significant proportion of the total variation in the DV ACT Science was predicted by the collection of IVs (R^2 = .506; F[16, 287] = 18.407, p < .001). Together, the predictors accounted for 50.6% of the variance in the ACT Science scores. However, the fourth block of predictors (i.e., TEOP scores) once added to the model did not produce a significant change in the *F* statistic (ΔR^2 = .000; *p* =.967).

The model contained 16 predictors with the TEOP variables as the main IVs (i.e., Action and Sound). Of the 16 IVs in the model, from the first block, one demographic/behavioral variable, Gender was a significant predictor of one's ACT Science score. In the second block, Study/Testing Variables one of the three IVs was a significant predictor of the outcome variable. This variable was if the student feels that he or she is a good test taker. In the third block all four of the high school academic data variables (i.e., GPA, number of ACT attempts, AP/Honors classes, and if their favorite class was a tested ACT subject) were significant predictors of the DV. In the fourth block neither of the TEOP variables "Action" or "Sound" were significant predictors of ACT Science scores.

The following paragraph presents the statistical tests for each significant predictor in the model. From the first block, Gender was a statistically significant predictor of ACT Science (t = 3.196, df = 287, p = .002). Specifically, Gender was positively predictive of ACT Science (β = .146; B = 1.494, SE = .467). If a student is male, their ACT Science score will be 1.494 points higher than a female student.

From the second block, the variable "Good Test Taker" (t = 4.198, df = 287, $p <$ .001) was a statistically significant predictor of ACT Science score. "Good Test Taker" was positively related (β = .209; B = 2.153, SE = .513) to ACT Science results. For this variable, if a student feels that he/she is a good test taker, their ACT Science was predicted to be 2.153 points higher than those who do not feel that they are good test takers.

From the third block, GPA (t = 6.346, *df* = 287, *p* < .001) was a statistically significant predictor of ACT Science score. Specifically, GPA was positively predictive (β = .398; *B* = 2.936, *SE* = .463) of the outcome. For every additional GPA point, students' predicted ACT Science scores will increase by 2.936 points. The number of ACT attempts (t = 2.451, *df* = 287, *p* = .015) was a statistically significant predictor (β = .118; *B* = 1.220, *SE* = .498) and was positively predictive of ACT Science scores. Students who took the ACT more than scored 1.220 points higher than those who were taking the test for the first time. The variable representing AP or Honors course enrollment (t = 3.721, *df* = 287, *p* < .001) was a statistically significant and positive predictor of ACT Science scores. If students took an AP or Honors course (β = .204; *B* = 2.134, *SE* = .574) he/she would have a 2.134 point higher score on the ACT Science (i.e., compared to those who have not taken an AP or Honors course). The variable representing if a students' favorite school subject was one of the ACT tested areas (t = 2.437, *df* = 287, *p* = .015) was a statistically significant predictor of ACT Science scores. If student's favorite course was on the ACT (β = .110; *B* = 1.151, *SE* = .472) his/her ACT Science scores are predicted to be 1.151 points higher than those whose favorite course was not on the ACT. For the main variables of interest in the model both of the TEOP variables "Action" (*t* = -.191, *df* = 287, *p* = .849) (β = -.009; *B* = -.048, *SE* = .252), and "Sound" (*t* = -.088, *df* = 287, *p* = .930) (β = -.004; *B* = -.025, *SE* = .258) were not significantly predictive of the DV ACT Science. A summary of these results can be found in Table 23.

Table 23

Main Study Summary of Hierarchical Multiple Linear Regression Analyses ACT Science (N = 304)

Independent Variables	*B*	*SE*	*t*	*β*	95% CI for *B*	
					Lower	Upper
Demographic/Behavioral						
Gender	1.494	.467	3.196**	.146	.574	2.414
Ethnicity	-.303	.645	-.469	-.021	-1.573	.968
High School Type	-.072	.506	-.142	-.007	-1.068	.925
Mother's Education	.491	.503	.967	.047	-.499	1.482
Father's Education	.570	.496	1.149	.055	-.407	1.547
# of Extra-Curricular	-.218	.156	-1.394	-.072	-.524	.089
Exercise	-.466	.514	-.907	-.040	-1.477	.545
Study/Testing						
Good Test Taker	2.153	.513	4.198***	.209	1.144	3.163
Focus when Testing	.070	.509	.137	.007	-.932	1.071
Preferred Time Congruency	-.602	.508	-1.186	-.053	-1.602	.397
High School Academic						
GPA	2.936	.463	6.346***	.398	2.025	3.847
ACT Attempts	1.220	.498	2.451*	.118	.240	2.200
AP/Honors	2.134	.574	3.721***	.204	1.005	3.263
Favorite Class	1.151	.472	2.437*	.110	.221	2.081
TEOP						
Action	-.048	.252	-.191	-.009	-.544	.448
Sound	-.025	.258	-.088	-.004	-.587	.537

Note. *$p < .05$, **$p < .01$, ***$p < .001$. SS= Social Studies, PE = Physical Education, MA = Math, SC = Science, Eng = English. *B* = unstandardized regression coefficient, *SE* = standard error, β = standardized regression coefficients, CI = confidence interval

ACT English

A Hierarchical Multiple Linear Regression was conducted that examined the relationship between the TEOP and ACT English, controlling for demographic/behavioral, study/testing, and high school academic variables. The results suggest that a significant proportion of the total variation in the DV ACT English was predicted by the collection of IVs ($R^2 = .616$; $F[16, 287] = 28.794$, $p < .001$). Together, the predictors accounted for 61.6% of the variance in the ACT English score. However, the fourth block of predictors (i.e., TEOP Scores) once added to the model did not produce a significant change in the *F* statistic ($\Delta R^2 = .000$; $p = .981$).

The model contained 16 predictors with the TEOP variables as the main IVs (i.e., Action and Sound). Of the 16 IVs in the model, from the first block, two of the seven demographic/behavioral variables, Father's Level of Education and if the participant exercised were significant predictors of one's ACT English score. In the second block, Study/Testing Variables one of the three IVs was a significant predictor of the outcome variable. This variable was if a student feels he/she a good test taker. In the third block three of the four of the high school academic data variables GPA, number of ACT Attempts, and if they take AP/Honors classes, were significant predictors of the outcome. In the fourth block neither of the TEOP variables "Action" or "Sound" were significant predictors of ACT English scores.

The following paragraph presents statistical tests for each significant predictor in the model. From the first block, Father's Level of Education was a statistically significant predictor of ACT English ($t = 3.854$, $df = 287$, $p < .001$). The student's father's education level was positively predictive of ACT English ($\beta = .164$; $B = 2.159$, $SE = .560$). If a student's father had at least a bachelor's degree, their ACT English score was predicted to be 2.159 points higher (compared to those with fathers who have lower levels of education. The variable representing if a student exercised or not was a statistically significant and negative predictor (t = -2.487, df = 287, p = .013) of ACT English scores (β = -.097; B = -1.441, SE = .579). Students who did not exercise were predicted to score 1.441 points lower on the ACT English (compared to those who did).

From the second block, the variable "Good Test Taker" (t = 4.732, df = 287, p < .001) was a statistically significant predictor of ACT English score, and was positively related to (β = .208; B = 2.739, SE = .579) the outcome. For this variable, if a student feels they are a good test taker, his/her ACT English score will be 2.739 points higher than those who do not feel that they are good test takers.

From the third block, GPA (t = 6.663, df = 287, p < .001) was a statistically significant predictor of ACT English scores. Specifically, GPA was positively predictive (β = .368; B = 3.478, SE = .579) of the outcome. For every additional GPA point, students' predicted ACT English scores will increase by 3.478 points. The number of ACT attempts (t = 3.740, df = 287, p < .001) was a statistically significant predictor and was positively predictive of ACT English scores (β = .159; B = 2.101, SE = .562). Students who took the ACT more than once scored 2.101 points higher than those who were taking the test for the first time. The variable representing AP or Honors course enrollment (t = 3.861, df = 287, p < .001) was a statistically significant and positive predictor of ACT English scores. If students took an AP or Honors (β = .187; B = 2.499, SE = .647) he/she would have a 3.861 point higher score on the ACT English (i.e., compared to those who have not taken an AP or Honors course). For the main variables of interest in the model, both of the TEOP variables "Action" (t = -.140, df = 287, p = .889) (β = -.006; B = -.040, SE = .284), and "Sound" (t = -.073, df = 287, p = .942) (β =

-.003; *B* = -.024, *SE* = .322) were not significantly predictive of the DV ACT English. A summary of these results can be found in Table 24.

Table 24

Main Study Summary of Hierarchical Multiple Linear Regression Analyses ACT English (N = 304)

Independent Variables	*B*	*SE*	*t*	*β*	95% CI for *B* Lower	Upper
Demographic/Behavioral						
Gender	-.131	.527	-.249	-.010	-1.169	.907
Ethnicity	.162	.728	.222	.009	-1.271	1.595
High School Type	1.111	.571	1.945	.081	-.013	2.236
Mother's Education	.155	.568	.274	.012	-.962	1.273
Father's Education	2.159	.560	3.854^{***}	.164	1.057	3.262
# of Extra-Curricular	-.060	.176	-.342	-.016	-.406	.286
Exercise	-1.441	.579	-2.487^{*}	-.097	-2.581	.300
Study/Testing						
Good Test Taker	2.739	.579	4.732^{***}	.208	1.600	3.878
Focus when Testing	-.493	.574	-.858	-.037	-1.632	.637
Preferred Time Congruency	.535	.573	.934	.037	-.593	1.663
High School Academic						
GPA	3.478	.522	6.663^{***}	.368	2.451	4.506
ACT Attempts	2.101	.562	3.740^{***}	.159	.995	3.207
AP/Honors	2.499	.647	3.861^{***}	.187	1.225	3.773
Favorite Class	.930	.533	1.745	.069	-.119	1.979
TEOP						
Action	-.040	.284	-.140	-.006	-.599	.520
Sound	-.024	.322	-.073	-.003	-.657	.610

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. SS= Social Studies, PE = Physical Education, MA = Math, SC = Science, Eng = English. *B* = unstandardized regression coefficient, *SE* = standard error, β = standardized regression coefficients, CI = confidence interval

ACT Reading

A Hierarchical Multiple Linear Regression was conducted that examined the relationship between the TEOP and ACT Reading, controlling for demographic/behavioral, study/testing, and high school academic variables. The results suggest that a significant proportion of the total variation in the dependent variable ACT reading was predicted by the collection of IVs ($R^2 = .477$; $F[16, 287] = 16.359$, $p < .001$). Together, the predictors accounted for 47.7% of the variance in the ACT Reading scores. However, the fourth block of predictors (i.e., TEOP Scores) once added to the model did not produce a significant change in the *F* statistic ($\Delta R^2 = .002$; $p = .608$).

The model contained 16 predictors with the TEOP variables as the main IVs (i.e., Action and Sound). Of the 16 IVs in the model, from the first block, one of the seven demographic/behavioral variables, Father's Level of Education was significant predictor of one's ACT Reading score. In the second block, Study/Testing Variables one of the three IVs was a significant predictor of the outcome variable. This was variable if the student feels he/she is a good test taker. In the third block two of the four of the high school academic data variables, GPA, and AP/Honors class enrollment, were significant predictors of the outcome. In the fourth block neither of the TEOP variables "Action" or "Sound" were significant predictors of ACT Reading scores.

The following paragraph presents the statistical tests for each significant predictor in the model. From the first block, Father's Level of Education was a statistically significant predictor of ACT Reading ($t = 3.057$, $df = 287$, $p = .002$). Specifically, a student's the father's education level was positively predictive of ACT Reading ($\beta = .152$; $B = 1.952$, $SE = .639$). If a student's father had at least a bachelor's degree, their ACT Reading score was predicted to be 1.952 points higher than those whose fathers have lower levels of education.

From the second block, the variable "Good Test Taker" (t = 4.370, *df* = 287, $p < .001$) was a statistically significant predictor of ACT Reading score and was positively related to the ($\beta = .224$; *B* = 2.883, *SE* = .660) ACT Reading scores. For this variable, if a student feels he/she is a good test taker, his/her ACT Reading score was predicted to be 2.883 points higher than those who do not feel that they are good test takers.

From the third block, GPA (t = 5.036, *df* = 287, $p < .001$) was a statistically significant predictor of ACT Reading score. Specifically, GPA was positively predictive ($\beta = .325$; *B* = 2.997, *SE* = .595) of the outcome. For every additional GPA point, the students' predicted ACT Reading scores will increase by 2.997 points. The variable representing AP or Honors course enrollment (t = 3.526, *df* = 287, $p < .001$) was a statistically significant and positive predictor of ACT Reading scores. If students took an AP or Honors course ($\beta = .199$; *B* = 2.601, *SE* = .738) he/she would have a 2.601 point higher score on the ACT Reading (compared to those who did not). For the main variables of interest in the model both of the TEOP variables "Action" (*t* = -.041, *df* = 287, *p* = .967) ($\beta =$ -.002; *B* = -.013, *SE* = .324), and "Sound" (*t* = .938, *df* = 287, *p* = .349) ($\beta = .049$; *B* = .344, *SE* = .367) were not significantly predictive of the DV ACT Reading. A summary of these results can be found in Table 25.

Table 25

Main Study Summary of Hierarchical Multiple Linear Regression Analyses ACT Reading (N = 304)

Independent Variables	*B*	*SE*	*t*	*β*	95% CI for *B* Lower	Upper
Demographic/Behavioral						
Gender	-.547	.601	-.909	-.043	-1.730	.637
Ethnicity	-.009	.830	-.011	-.001	-1.643	1.624
High School Type	.538	.651	.826	.040	.744	1.820
Mother's Education	.455	.647	.702	.035	-.819	1.728
Father's Education	1.952	.639	3.507^{**}	.152	.695	3.208
# of Extra-Curricular	-.001	.200	-.007	.000	-.395	.393
Exercise	-.625	.661	-.947	-.043	-1.925	.675
Study/Testing						
Good Test Taker	2.883	.660	4.730^{***}	.224	1.585	4.182
Focus when Testing	-.315	.654	-.481	-.024	-1.603	.973
Preferred Time Congruency	-.429	.653	-.657	-.030	-1.715	.856
High School Academic						
GPA	2.997	.595	5.036^{***}	.325	1.825	4.168
ACT Attempts	.657	.640	1.027	.051	-.063	1.918
AP/Honors	2.601	.738	3.526^{***}	.199	1.149	4.053
Favorite Class	.790	.608	1.300	.060	-.406	1.986
TEOP						
Action	-.013	.324	-.041	-.002	-.651	.624
Sound	.344	.367	.938	.049	-.378	1.066

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. SS= Social Studies, PE = Physical Education, MA = Math, SC = Science, Eng = English. *B* = unstandardized regression coefficient, *SE* = standard error, β = standardized regression coefficients, CI = confidence interval

Summary of the Hierarchical Multiple Linear Regressions

Each of regression models was able to predict a significant proportion of the total variance for all five dependent variables (i.e., ACT Comprehensive, Math, Science, English, and Reading). The IVs GPA and if the student is a good test taker were statistically significant predictors of all five outcome variables. GPA was most influential IV in predicting all five DVs in the regression models. Conceptually, these results are unsurprising, as GPA measures how academically successful a student is in school and the ACT assesses how prepared a student is for college. Furthermore, students most likely believe that they are good test takers because they have had previous success with testing. Thus, these participants may be able to perform well on the ACT because they have been successful on assessments in the past. These results indicate that in-school academic achievement (or lack thereof) is predictive of high-stakes standardized test scores.

The two main IVs, the “Action” and “Sound” TEOP variables, were not influential in predicting any of the DVs. The two TEOP variables resulted in no changes in R^2 for ACT Comprehensive, Math, Science, and English. When “Action” and “Sound” were added to the model for Reading, they only resulted in a non-statistically significant increase in explained variation. The High School Academic block exerted less impact on the variability of DV ACT Reading. Only two of the four IVs within the block were significant compared to three or all four in the other models. Due to GPA’s dominating influence in explaining the variation in the ACT DVs, a post hoc multiple regression analysis was run using GPA as the DV, to ascertain if the TEOP factors exert influence on overall high school academic performance.

Post Hoc Analyses

Because of the strong sway of GPA on ACT Scores across all categories of the RQ2 regression analyses, a post hoc Hierarchical Multiple Linear Regression was conducted to address the following research question, “What is the relationship between the TEOP factor scores (i.e., Action and Sound) and cumulative Grade Point Average (GPA) in a high school student population?” One Hierarchical Multiple Linear

Regression was run, using GPA as the DV. Thirteen IVs were used in the model and entered in the following blocks: (1) Demographics (i.e., Gender, Ethnicity, High School Type, Mother's Education Level, Father's Education Level, Number of Extra-Curricular Activities, if they Exercise, and if the take AP/Honors courses), (2) Study/Testing (i.e., Are they a Good Test Taker, do they have Trouble Focusing when Testing, and is their preferred Test Time in the morning), and (3) TEOP Scores (i.e., "Action" and "Sound").

Data Cleaning

In addition to the eliminated cases prior to RQ1, mentioned previously, cases were also excluded in the post hoc final analysis sample. There were 47 additional cases removed from Post Hoc RQ who had missing data on an IV. The final analysis sample for post hoc RQ was 304.

In the post hoc analysis sample (N = 304), 146 (48.0%) were male and 158 (52.0%) were female. The sample was comprised of 258 (84.9%) White/Caucasian participants. A majority of the participants (N = 195, 64.1%) attended a private high school. A Bachelor's degree or higher was earned by 182 (59.9%) of the participants' mothers and 164 (53.9%) of their fathers. The mean number of extra-curricular activities the sample participated in was 2.32 (SD = 1.73). Two-hundred twenty three (73.4%) of the respondents stated that they exercised. One- hundred sixty seven (54.9%) of the respondents believe they are "good test takers." A majority of participants 170 (55.9%) feel they have trouble staying focused when testing. Two-hundred eighteen (71.7%) of the respondents preferred taking a test at a time that was not congruent with a morning test time. One-hundred eighty one (59.5%) of participants has taken at least one AP or honors class. The mean score for the “Action” factor was 3.46 (*SD* = .97) and 2.30 (*SD* = .92) for the “Sound” factor. The mean GPA for the sample was 3.18 (*SD* = .70).

Table 26

Post hoc Regression Demographics GPA (N = 304)

Variables	*M*(*SD*) or *n*(%)	Coding
Gender		
Female	158(52.0)	0
Male	146(48.0)	1
Ethnicity		
Other	46(15.1)	0
White	258(84.9)	1
High School Type		
Public	109(35.9)	0
Private	195(64.1)	1
Mother's Education		
< Bachelors	122 (40.1)	0
≥ Bachelors	182(59.9)	1
Father's Education		
< Bachelors	140 (46.1)	0
≥ Bachelors	164(53.9)	1
Number of ECAs	2.32(1.73)	--
Exercise		
No	81(26.6)	0
Yes	223(73.4)	1
AP Honors		
No	123 (40.5)	0
Yes	181(59.5)	1
Good Test Taker		
No	137 (45.1)	0
Yes	167 (54.9)	1
Trouble Focusing on Test		
No	134 (44.1)	0
Yes	170 (55.9)	1
Test Time Congruency		
No	218 (71.7)	0
Yes	86 (28.3)	1
Action	3.46 (.97)	--
Sound	2.30 (.92)	--
GPA (4.0 Scale)	3.18 (.70)	--

Note. ECA = Extra-Curricular Activities; SS = Social Studies; PE = Physical Education; MA = Math; SC = Science; Eng = English.

Outliers and Assumptions - GPA

For the Hierarchical Multiple Regression with GPA as the DV, centered leverage, Cook's Distance (*D*), and Mahalanobis Distance values were examined. The centered leverage values did not suggest any problematic data (Lomax & Hahs-Vaughn, 2012). The minimum centered leverage value was .017 and the maximum was .085 with the average at .043. Additionally, the centered leverage values were less than .2, indicating no extreme scores for any of the variables. Cook's *D* indicated that all values were close to zero with the minimum was zero and the maximum at .041, (i.e., the average was .004).

Mahalanobis Distance was consulted, and with an alpha level of .001, the critical value was 27.688. The maximum value was 25.943, with an average of 12.957. Thus, as no values were greater than 27.688, there was no evidence to suggest that there were any multivariate outliers in the model.

Before examining the GPA regression model, the basic assumptions of Multiple Linear Regression were investigated. For independence, the scatterplots of studentized residuals depicted a random display of points between -2 and +2.

For Linearity, separate partial regression plots showed a random display of data points falling approximately between the boundaries of -2 and +2, indicating that the assumption was met. Homoscedasticity was examined with scatterplots, and the dispersion of the values around the regression line remained fairly constant for all values of X. For the assumption of Normality, the histograms of residuals appeared to be normally distributed indicating that, the assumption was met. Finally, for

Multicollinearity the correlation matrix showed no unusually high coefficients ($r \leq .80$). The tolerances for all the predictors were within acceptable limits with the VIFs corroborating this evidence. The collinearity diagnostics did not indicate any overlap in the contribution of the percentages of variance explained to the model.

Correlations

Correlations were run to examine the relationship between IVs and DVs. The highest positive correlation between IVs was between Mother's Education Level and Father's Education Level ($r_{\varphi} = .433$, $p < .001$) and the highest negative correlation was between the participant believing they are a "good test taker" and their ability to focus when testing ($r_{\varphi} = -.464$ $p < .001$). The highest positive correlation indicates that those whose mothers have attained higher levels of education are likely to have fathers with advanced levels of education as well. The highest negative correlation indicates that students with trouble focusing when testing also believe that they are "bad test takers."

Of the IVs, the strongest positive relationship with the DV GPA was if the student took AP or honors courses ($r_{pb} = .603$, $p < .001$) and the strongest negative relationship was the students TEOP "Sound" score for sound preference ($r = -.333$, $p < .001$). The highest positive correlation indicates that those who enrolled in more rigorous classes tend to have higher GPA's. The highest negative correlation indicates that students with a propensity to prefer more noise or auditory stimulation are inclined to have lower GPAs. The TEOP variable "Action" also had a significant negative correlation with GPA ($r = -.188$, $p < .001$) suggesting that as a student's preference for "Action" increased their GPA was reduced.

Table 27

Post hoc Regression Correlations (N = 304)

Items	1	2	3	4	5	6	7
1. Gender	--	.119*	.030	.054	.101	-.143**	.089
2. Ethnicity		--	-.165**	.111*	.090	.067	.082
3. HS Type			--	.157**	.208***	.204***	.199***
4. Mom Ed				--	.433***	.259***	.175**
5. Dad Ed					--	.258***	.124*
6. ECA						--	.248***
7. Exercise							--
8. AP/Honors							
9. GTT							
10. Focus							
11. Time							
12. Action							
13. Sound							
14. GPA							

Note. *p<.05, **p<.01, ***p<.001; HS = High School; Ed = Education; ECA = Extra-Curricular Activities; GTT = Good Test Taker; Focus = Focus when Testing; Time = Test Time Congruency; GPA = Grade Point Average (4.0 Scale).

Table 27 (continued)

Post Hoc Regression Correlations (N = 304)

Items	8	9	10	11	12	13	14
1. Gender	.-.142**	.079	--.057	-.009	-.086	.061	-.124*
2. Ethnicity	.257***	.247***	-.123*	.078	-.052	-.045	.244***
3. HS Type	-.101	-.082	.023	.112*	-.149**	-.211***	.184**
4. Mom Ed	.299***	.080	-.148**	.016	-.104	-.142*	.330***
5. Dad Ed	.210***	.125*	-.143*	.023	-.049	-.165**	.278***
6. ECA	.385***	.106*	-.094	.186***	-.036	-.195***	.513***
7. Exercise	.157**	.068	-.036	.081	-.014	-.110*	.183**
8. AP/Honors	--	.273***	-.192***	.055	-.043	-.138*	.603***
9. GTT		--	-.464***	-.060	-.071	-.056	.311***
10. Focus			--	-.098	.175**	.178**	-.286***
11. Time				--	-.101	-.233***	.147**
12. Action					--	.394***	-.188***
13. Sound						--	-.333***
14. GPA							--

Note. *p<.05, **p<.01, ***p<.001; HS = High School; Ed = Education; ECA = Extra-Curricular Activities; GTT = Good Test Taker; Focus = Focus when Testing;
Time = Test Time Congruency; GPA = Grade Point Average (4.0 Scale).

Hierarchical Multiple Linear Regression

A Hierarchical Multiple Linear Regression analysis was run to determine if a relationship exists between TEOP scores (i.e., Action and Sound) and GPA controlling for demographic/behavioral, study/testing, and TEOP.

GPA

The results suggest that a significant proportion of the total variation in the DV GPA was predicted by the collection of IVs ($R^2 = .551$; $F[13, 291] = 27.525$, $p < .001$). Together, the predictors accounted for 55.1% of the variance in GPAs. The third and final block of predictors (i.e., TEOP Scores) once added to the model did produce a significant change in the *F* statistic ($\Delta R^2 = .030$; $p < .001$).

The model contained 13 predictors with the TEOP variables as the main IVs (i.e., Action and Sound). Of the 13 IVs in the post hoc model, from the first block, four of the eight demographic/behavioral variables were significant predictors of one's GPA. These were the variables ethnicity, if they attend a public or private school, the number of extra-curricular activities the student participates in, and if the student takes AP or Honors courses or not. In the second block, Study/Testing Variables one of the three was a significant predictor of the outcome variable. This was the independent variable student feels they are a good test taker. In the third block TEOP factor "Sound" was a significant predictor of GPA.

The following paragraph presents the statistical tests for each significant predictor in the post hoc model. From the first block, Ethnicity was a statistically significant predictor of GPA ($t = 2.482$, $df = 291$, $p = .014$). Specifically, Ethnicity was positively predictive of GPA (β = .104; *B* = .203, *SE* = .082). If a student was White/Caucasian their GPA was predicted to be .203 higher than the GPA of other races/ethnicity. The IV High School type, either public or private was a significant and positively related to GPA (t = 3.398, df = 291, p = .001), with students who attend private schools (β = .149; *B* = .216, *SE* = .064), predicted to have a .216 higher GPA than those students who attend public schools. The number of extra-curricular activities a student participates in is a statistically significant and positive predictor (t = 5.199, df = 291, $p < .001$) of GPA (β = .241; *B* = .099, *SE* = .019), with students receiving a .099 higher GPA for each additional extra-curricular activity. The variable representing AP or Honors course enrollment (t = 8.609, df = 291, $p < .001$) was a statistically significant and positive predictor of GPA. If students took AP or Honors courses (β = .400; *B* = .567, *SE* = .066) he/she would have a .567 point higher GPA (i.e., compared to those who have not taken an AP or Honors course).

From the second block, the variable “Good Test Taker” (t = 42.439, df = 291, p = .015) was a statistically significant predictor of GPA, and was positively related (β = .113; *B* =.159, *SE* = .065) to the outcome. For this variable, if a student feels that he/she is a good test taker, his/her predicted GPA will .159 higher than those who do not feel that they are good test takers. From the third block, TEOP factor “Sound” (t = -3.717, df = 291, $p < .001$) was a statistically significant and negative predictor (β = -.174; *B* = -.132, *SE* = .036) of GPA. As a person’s preference for sound increases, their GPA will decrease by .132 points. TEOP variable “Action” ” (t = -.672, df = 291, p = .502) was not a significant predictor (β = -.021; B = -.030, SE = .032) of GPA. A summary of these results can be found in Table 28.

Table 28

Post hoc Summary of Hierarchical Multiple Linear Regression Analysis GPA (N = 304)

Independent Variables	B	SE	t	β	95% CI for B	
					Lower	Upper
Demographic/Behavioral						
Gender	-.098	.058	1.674	-.070	.213	.017
Ethnicity	.203	.082	2.482*	.104	.042	.364
High School Type	.216	.064	3.398**	.149	.091	.342
Mother's Education	.119	.064	1.850	.084	-.008	.246
Father's Education	-.009	.064	-.136	-.006	-.134	.117
# of Extra-Curricular	.099	.019	5.199***	.241	.061	.136
Exercise	.022	.066	.335	.014	-.107	.152
AP Honors	.567	.066	8.609***	.400	.437	.697
Study/Testing						
Good Test Taker	.159	.065	2.439*	.113	.031	.287
Focus when Testing	-.070	.065	-1.078	-.050	-.197	.058
Preferred Time Congruency	-.067	.065	-1.033	-.043	-.195	.061
TEOP						
Action	-.021	.032	-.672	-.030	-.084	.041
Sound	-.132	.036	-3.717***	-.174	-.203	-.062

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. SS= Social Studies, PE = Physical Education, MA = Math, SC = Science, Eng = English. B = unstandardized regression coefficient, SE = standard error, β = standardized regression coefficients, CI = confidence interval

Summary of Post Hoc Multiple Regression

The post hoc regression model was able to predict a significant proportion of the total variance in the DV GPA. The IV of AP or Honors course enrollment was the most influential IV in predicting GPA. Conceptually, GPA is a measure of how successful a student is in school and students who take AP or Honors courses are frequently among the top academic performers. The two TEOP variables, "Action" and "Sound", when added to the model were statistically influential in predicting GPA. The two TEOP variables resulted in a significant increase in explained variation for the DV GPA. Moreover, of the IVs, “Sound” had the largest negative correlation with GPA, indicating that those with higher preferences for auditory stimulation are predicted to have lower GPAs.

Summary of All Results

The aims of Chapter IV were to investigate the psychometric properties (i.e., reliability and validity) of the TEOP and to explore the relationship between TEOP scores and high-stakes standardized test performance. The main study results indicated that the TEOP was psychometrically supported; however, the dimensionality of the measure used to predict standardized test performance was inconclusive. The TEOP variable “Action” was negatively correlated with four of the five DVs, and “Sound” was negatively related to all five outcome variables. However, when added to the regression models, the two TEOP variables did not have a significant increase in explained variation. The IV GPA was found to have a dominating influence on the explained variation in ACT scores.

A post hoc test using GPA as the DV was conducted to examine if the TEOP variables have a predictive relationship with overall school performance and not just “one moment in time” assessments such as the ACT. The analysis did provide evidence that the TEOP factors, in particular “Sound,” are predictive of school academic performance. Both “Action” and “Sound” had significant negative correlations with GPA, and when added to the model provided a significant increase in explained variability. The following chapter (Chapter V: Discussion) discusses these findings and provides some implications from the results. This chapter also presents limitations and possibilities for future research.

CHAPTER V (MAIN STUDY): DISCUSSION

Discussion (Main Study)

The objective of this study was to examine current high school students' perceptions of preferred standardized testing conditions and if individual preferred conditions had a predictive relationship with standardized test results (i.e., the ACT). The two primary research questions that guided the study were: (RQ1) "What are the psychometric properties (i.e., construct validity and internal consistency reliability) of the scores on the Test Environment for Optimal Performance (TEOP) in a high school student population?", and (RQ2) "What is the relationship between the TEOP factor scores (i.e., Action and Sound) and high-stakes aptitude test scores (i.e., the ACT) in a high school student population?". In addition, after considering the results a post-hoc research question was added to the study (PH1) What is the relationship between the TEOP factor scores (i.e., Action and Sound) and cumulative Grade Point Average (GPA) in a high school student population?"

Confirmatory Factor Analysis (CFA)

A CFA was conducted to test if the data fit the hypothesized factor structure. After consulting the modification indices, the model included the observed variables of water, food, gum, stand, and walk all loading significantly on the "Action" factor. Item 6 (i.e., "I prefer to have food available.") had the strongest loading and the "Action" factor explained the largest proportion of variance in this item. This can be linked with the measured steps required to bring food to the testing location. Additionally, Item 8 (i.e., "I prefer to have the option to stand up.") had the weakest loading and the "Action" factor explained the least amount of variance in this item. Item 8's (i.e., "I prefer to have the option to stand up.") loading decreased considerably from the initial model to the final model. This reduction can be attributed to the added error covariance between Items 8 (i.e., "I prefer to have the option to stand up.") and 9 (i.e., I prefer to have the option to walk around.").

The observed variables of silence, noise, music, and TV all loaded significantly on the factor "Sound." Item 3 (i.e., "I prefer to listen to music.") had the strongest loading and the "Sound" factor explained the largest proportion of variance in this item. The strong loading can be explained by the undeviating relationship between music and sound. Item 4 (i.e., "I prefer to have television on in the background.") had the weakest loading and the "Sound" factor explained the smallest proportion of variance in this item. The low loading could be attributed to television being perceived as a visual medium more so than sound.

The largest score from the TEOP "Action" items in the CFA sample was Item 5 (i.e., "I prefer to have water/a beverage available.") and the lowest score was on Item 9 (i.e., "I prefer to have the option to walk around."). As indicated in the Pilot Study section above, the high preference for being able to drink indicates that test takers would like the choice of having a beverage during testing to conceivably help with concentration, calm anxiety, provide a break to think, and quench thirst. The lower preference for walking around may be due to the timed nature of these exams. Students may be less motivated to consider walking around during a test, as this would involve moving away from their desk/computer station where their exam is located. Students may be apprehensive or fretful that walking around will prevent them from completing the assessment on time. Finally, the highest mean score for "Sound" was Item 1 (i.e., "I prefer silence.") and the lowest was Item 4 (i.e., "I prefer to have a television on in the background."). As with the Pilot Study sample, most participants prefer silence, and few favor noisy appliances/devices.

The inter-item correlations within the "Action" factor (i.e., Items 5 to 9) were all significant and positive in direction. The same was noted for Items 1 through 4 in the "Sound" factor. Additionally, stronger relationships were found among items within the same factor compared to items located in the other factor. The strongest correlation among the "Action" items was between Item 8 (i.e., "I prefer to have the option to stand up.") and Item 9 (i.e., "I prefer to have the option to walk around."). Similarly for the "Sound" factor, Item 1 (i.e., "I prefer silence.") and Item 3 (i.e., "I prefer to listen to music.") had the strongest correlation. The two-factor TEOP structure was confirmed. The two-factor structure of the TEOP used in the CFA was confirmed and produced evidence of high internal consistency reliability. The five-item "Action" factor had a higher reliability compared to the four-item "Sound" factor. There were positive correlations between the two subscales (i.e., "Sound" and "Action"). This slight correlation indicates a relationship between "Action" and "Sound" scores, though it is not statistically significant.

Summary of Confirmatory Factor Analysis

Further examination of the item correlation matrix revealed a near "archetypal" example of the within and between patterns of association that should be extracted after following best practices in construct-based measure construction. For "Action," all the items contained within that factor had significant, positive, and high correlations with each other. These items were also not as strongly correlated with the items in the "Sound" factor. The correlations between items located in different factors (e.g., a correlation coefficient between an item in the "Action" factor and an item in the "Sound" factor) were inconsistent compared to the correlations amongst the items contained within each factor (e.g., a correlation coefficient between two items within the "Action" factor or between two items within the "Sound" factor). That is, there was less variation in the within-factor item correlations compared to the variation in the between-factor item correlations. The correlation patterns followed the expected model of between and within group relationships.

Hierarchical Multiple Linear Regressions

Five Hierarchical Multiple Linear Regression analyses were run to establish if TEOP scores are a significant predictor of ACT Comprehensive scores as well as the four subject-specific sections of the exam (i.e., Math, Science, English, and Reading) controlling for various Demographic, Studying/Testing, and Academic variables in a high school student sample.

ACT Comprehensive

Of the 16 variables included in the ACT Comprehensive model, just under half were statistically significant with the overall collection of IVs predictive of ACT Comprehensive scores. The first three groupings of variables (i.e., Demographic, Studying/Testing, and Academic covariates) produced a significant increase in the proportion of variance explained in the outcome, but there was not a significant increase when the fourth block of predictors (i.e., TEOP factors) were added. Of the seven significant IVs, gender (i.e., male) and father's education level (i.e., above a Bachelor's degree) were significant positive predictors of higher ACT Comprehensive scores from the Demographic Block. In the second block, students' self-perception of being a "good test taker" was positively predictive of higher ACT Comprehensive scores from the Studying/Testing Block. Finally, all four IVs in the Academic Block (i.e., GPA, AP Honors Courses, number of times the student took the ACT exam, and favorite class) were positively related to the overall ACT exam scores.

In this regression model, cumulative GPA had the strongest positive relationship to ACT Comprehensive scores, and explained the largest percentage of variance in that outcome. This means that as students' GPAs increase, their scores on the ACT exam are likely to be higher. That is, the students in this sample who were more successful in their high school coursework performed better on the ACT Comprehensive than their peers with lower GPAs. Along with GPA, the other three variables (i.e., number of ACT Attempts, AP/Honors classes, and if their favorite class was a tested ACT subject) from the Academic Block all had a positive and significant relationship with students' ACT

Comprehensive scores. This provides evidence of the connection between what a student does during the school day and how they perform on high-stakes standardized exams. Adding the fourth block of variables (i.e., the TEOP factors) did not significantly impact the outcome. Neither of the TEOP variables (i.e., Action or Sound) were significant predictors of the amount of variance in the ACT Comprehensive scores. In other words, ACT Comprehensive scores were not influenced by a preference for physical movement or auditory stimulation.

Based on these results, ACT Comprehensive scores are connected to more academically-related variables, specifically GPA. Students that earn higher grades, take more rigorous classes, attempt the ACT exam multiple times, and whose favorite subject is one of the four ACT subject tests will earn higher ACT Comprehensive scores. This is compared to their peers with lower grades, who are only enrolled in basic classes, and who have only attempted the ACT exam one time or less with their preferred class being a subject not specifically tested on the ACT who have lowers scores on the ACT Comprehensive exam. Additionally the TEOP variables were not significantly predictive of ACT Comprehensive, indicating that a preference for “Action” and “Sound” during standardized tests have minimal to no impact on one-time ACT tests results.

ACT Subjects

Sixteen variables were incorporated in the ACT Subjects (i.e., Math, Science, English, and Reading) models. Five IVs were significant predictors of DV Math, six for Science, six for English, and four IVs were significant predictors of Reading. The TEOP variables were not statistically significant predictors of the DV in any of the four analyses.

In all four models the variables as a whole were predictive of the DV score for the respective subject matter test (i.e., Math, Science, English, and Reading). Each of the four Multiple Linear Regression analyses (i.e., Demographic, Study/Testing, and Academic Covariates) produced a significant increase in the proportion of variance accounted for in the outcome, but in each instance, there was not a significant increase when the fourth block of predictors (i.e., TEOP factors) was added.

The IV cumulative GPA had the strongest positive relationship to the DV for all four ACT subjects (i.e., Math, Science, English, and Reading), and explained the largest percentage of variance for each outcome. This suggests that as students' GPA's increase their scores on the ACT subject matter tests (i.e., Math, Science, English, and Reading) are likely to be higher. That is the students in this sample who were more successful in their high school course work performed better on the ACT subject matter tests (i.e., Math, Science, English, and Reading) than their peers with lower GPAs.

Along with GPA, two other IVs were significantly predictive of all four DVs. These are if a student felt they were a good test taker and if a participant was enrolled in AP or Honors course or not were both positive and statistically significant predictors on each of the four ACT subject DVs (i.e., Math, Science, English, and Reading). This advocates that students from the sample who are confident in their test taking ability scored higher on the four DVs than those less assured in their test taking ability. Additionally, those who take more rigorous classes in high school scored higher on each of the ACT subjects than their peers who did not take AP/Honors courses. The large amount of variance in the four outcomes explained by the GPA, if the student feels they are a good test taker or not, and AP/Honors course enrollment propose that ACT success can be predicted by a student's high school achievement, the student's confidence in their ability to take tests, and how ambitiously they schedule their courses.

This provides evidence of strong positive relationships between what a student does during the school day (i.e., GPA), what they do prior to the school year (i.e., schedule AP/Honors courses or not), and how they have performed on past tests (i.e., student feels they are a good test taker or not) with their performance on high-stakes standardized exams. In all four analyses adding the fourth block of variables (i.e., TEOP factors) did not significantly impact the outcome. Neither of the TEOP variables (i.e., Action or Sound) were significant predictors of the amount of variance in the ACT Comprehensive scores. In other words, ACT subject matter (i.e., Math, Science, English, and Reading) scores were not influenced by their preference for physical movement or auditory stimulation.

Based on these results, ACT subject matter (i.e., Math, Science, English, and Reading) scores appear to be connected to several variables, specifically if the student feels they are a good test taker or not, GPA, and AP/Honors course enrollment. Students that feel they are good test takers, earn higher grades, and take more demanding classes, will earn higher ACT subject matter outcomes. This is compared to their peers who believe they are not good test takers, receive lower grades, and are enrolled only in basic classes who have lower scores on the ACT subject matter tests (i.e., Math, Science, English, and Reading). Additionally, the TEOP variables were not significantly predictive on any of the DVs, indicating that preference for “Action” and “Sound” preferences during standardized tests have minimal to no impact on one time ACT tests results.

Table 29

Table of Hierarchical Multiple Linear Regression Analyses with the ACT

Model Information	ACT Test				
	Comprehensive	Math	Science	English	Reading
Variance	63.6%	57.9%	50.6%	61.6%	47.7%
Significant Blocks					
1	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
2	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
3	$p < .001$	$p < .001$	$P < .001$	$p < .001$	$p <.001$
4	--	--	--	--	--
Significant Variables by Block					
1	Gender Dad Ed.	Gender	Gender	Dad Ed. Exercise	Dad Ed.
2	Good Test Taker GPA	Good Test Taker GPA	Good Test Taker GPA	Good Test Taker GPA	Good Test Taker GPA
3	ACT Attempts AP/Honors Favorite Class	ACT Attempts AP/Honors Favorite Class	ACT Attempts AP/Honors Favorite Class	ACT Attempts AP/Honors	AP/Honors
4	--	--	--	--	--
Magnitude (*B*) of Significant Variables by Block					
1	.927 1.258	2.560	1.494	2.159 -1.441	1.952
2	2.472 3.138	1.694 3.202	2.153 2.936	2.739 3.478	2.883 2.997
3	1.127 2.602 1.219	.827 2.911 1.635	1.220 2.134 1.151	2.101 2.499	2.601
4	--	--	--	--	--

Note. Blocks: 1 = Demographic/Behavioral, 2 = Study/Testing, 3 = High School Academic, 4 = TEOP. *B* = Unstandardized Factor Loading.

Post Hoc Hierarchical Multiple Linear Regression

In all five Hierarchical Multiple Linear Regression analyses conducted to investigate RQ2, the IV GPA had a dominating impact on the ACT Comprehensive and the subject-matter tests. GPA was the most influential variable and explained the greatest amount of variance in the DV across all five models. The TEOP factors, when added to each of the five models, did not have a significant impact on any of the DVs (i.e., ACT Comprehensive, Math, Science, English, and Reading).

Of the 13 variables included in the GPA model, six were statistically significant; with, the variables as a whole predictive of the GPA. All three blocks of IVs (i.e., Demographic/ Behavioral, Study/Testing, and TEOP) produced a significant increase in the proportion of variance accounted for in the outcome including when the main IVs (i.e., TEOP factors) were added. Of the significant IVs, Ethnicity and High School Type (i.e., Public or Private), the number of extra-curricular activities the student participates in, and if they have enrolled in an AP or Honors course or not were significant predictors of GPA from the Demographic/ Behavioral block, with the student feeling they are a good test taker having a positive significant relationship from the Study/Testing block, as was the Sound factor from the TEOP block.

In this post hoc regression model, enrollment in AP or Honors classes had the strongest positive relationship to the GPA, and explained the largest percentage of the variance. This proposes that students who enroll in more rigorous high school courses will earn higher GPA's than their peers who do not take AP or Honors classes. The students from the sample who were more successful in their high school classes were also the ones who had more challenging academic schedules. Along with AP/Honors course enrollment the student's ethnicity, High School Type, and the participant feeling they are a good test taker all had a positive significant relationship with the student's GPA. From the third block of variables (i.e., TEOP Factors) "Sound" has a significant negative relationship with GPA. Sound was the only significant IV to have a negative impact on the outcome. That is as person's preference for sound increases there GPA decrease.

Based on these results, GPA outcomes appear to be associated with Demographic/ Behavioral, Study/Testing, and TEOP variables. Students ethnicity, high school type, participation in extra-curricular activities, enrollment in AP/Honors course or not, and feeling they are a good test taker or not, all have a positive predictive relationship with GPA. In addition there is a negative predictive relationship between "Sound" preference and the DV. In other words, unlike the ACT, which is a one moment in time assessment, participant's GPA was influenced by their preference "Sound." This suggests that the TEOP results can measure a person's individual testing preferences as related to who they are as a student and be predictive of overall school success despite not being predictive of standardized exam outcomes. TEOP scores are predictive of GPA because both are composites of many factors. This allows for the TEOP to be predictive general classroom exam and other assessment results.

The TEOP's predictive relationship of GPA can provide meaningful information. Classroom assessments comprise a much larger amount of one's academic career than standardized testing, and can account for up to one third of a student's time in school (Stiggins, 1991). Furthermore, individual classroom teachers are not constricted by the same rules as standardized testing companies and have a level of autonomy over classroom assessment structure. Research indicates that teachers value classroom assessments over standardized and state mandated tests, believing that in-class tests have the most influence on student learning (Allen, Ort, & Schimidt, 2009). Research on standardized testing policies in both public and private universities suggests GPA is better predictor of college success than standardized tests (i.e., ACT and SAT), as university GPA highly correlates with high school GPA (Hiss & Franks, 2014).

Because of the disparate frequency with which in-class assessments occur and the emphasis placed on classroom assessment practices related to student learning, conducing a post-hoc analysis with GPA as the outcome was necessary to further examine why the TEOP scores were not significantly predictive of ACT scores. Thus, a post hoc analysis was conducted to examine the predictive relationship of the TEOP scores and GPA as the DV to compare with the regressions where the DVs were the ACT Comprehensive and subject-matter test scores.

Initially, GPA was thought to only act as a control variable on standardized testing results, assuming that the TEOP would be the most influential predictors on ACT scores. The results from the main regressions suggest that GPA is highly predictive of ACT scores. GPA is hypothesized to be strongly predictive because it is a composite of many factors that are component of the ACT. GPA measures cognitive ability, content knowledge, and soft skills (i.e., work ethic, perseverance, and self-control), which are essential to academic success (National Education Association, 2017) and ACT performance.

Table 30

Table of Hierarchical Multiple Linear Regression Analyses GPA

Model Information	GPA
Variance	55.1%
Significant Blocks	
1	$p < .001$
2	$P = .004$
3	$p < .001$
Significant Variables by Block	
1	Ethnicity
	High School Type
	Number of Extra-Curricular Activities
	AP/Honors
2	Good Test Taker
3	Sound
Magnitude *B* of Significant Variables by Block	
1	.203
	.216
	.099
	.567
2	.159
3	-.132

Note. Blocks: 1 = Demographic/Behavioral, 2 = Study/Testing, 3 = TEOP. *B* = Unstandardized Factor Loading.

The Pilot Study and Main Study – A Summary

The objective of the Pilot Study discussed in Part I was to inspect the test environment preferences amongst current university students. The Pilot Study aimed to add to the narrow body of theoretical and empirical literature examining standardized testing environment conditions and test takers' conditional preferences. Additionally, Pilot Study Phase I used cross-validation methods to test the EFA and CFA models. Phase II used a Qualitative Interpretive design to provide further evidence to support the "Action" and "Sound" test environment preferences.

The current high schoolers in Generation Z have built upon the multitasking, instant gratification, and personalized existences of the preceding cohort and have essentially become Millennials on steroids. Members of Generation Z concurrently use multiple kinds of media, which has resulted in their being labeled with distributed and artificial attention spans (Tari, 2010). This means that Gen Zers give some attention to many things simultaneously while not actually focusing on anything. To probe these cohort differences, this study (i.e., both parts) was conducted as one of the first to address Generation Z's digital obsession and incompatibility with test environments.

The goal of the Main Study was to progress from the Pilot Study by supplying additional validation evidence for the TEOP scores in a population of high school students. The data were analyzed using CFA to examine the psychometric properties of the TEOP in a new population. Hierarchical Multiple Linear Regressions were conducted to provide evidence of the hypothesized relationships (i.e., Criterion-Related Validity – Concurrent) between the TEOP factors "Action" and "Sound" and aptitude outcomes (i.e., ACT scores). The conclusions drawn from the five Hierarchical Multiple Linear Regressions with ACT Comprehensive and the individual subject tests (i.e., Math, Science, English, and Reading) necessitated a post-hoc analysis using GPA as the DV.

Implications

From Part I, results from the Pilot Study highlight that Millennials have individual preferences for testing environment conditions. The conclusions from the Main Study extend these results to the Generation Z population. The findings from this study have implications for multiple levels of stakeholders and will be discussed in general and then from the proximal to the distal level in relationship to the student. The main study implications are addressed in the general section and the post hoc implications are outlined in the more specific proximal and distal sections. The Main Study produced additional evidence that the TEOP has valid and reliable scores for assessing test environment preferences in a current high school population. While TEOP results were not a significant predictor of high-stakes standardized test outcomes for this population, there are implications from this study that can be useful to a range of stakeholders.

General

The Main Study results suggest that test takers prefer a range of physical movement/activities during testing, from no action to engaging in multiple movements/activities. Furthermore, the results indicate that students also have a range of preferences for noises/sounds during testing, from complete silence to loud noises. The results provide evidence that high school students have some combination of action and sound preferences during standardized testing. People have individual inclinations for processing information, and making choices, with these preferences having an impact on education and learning (Kise, 2011). The knowledge gained from this study on Generation Z's test environment preferences can enhance the educational experience of students.

The most influential predictors of ACT scores across all five DVs (i.e., Comprehensive, Math, Science, English, and Reading) in this study were academically-related (i.e., GPA, ACT Attempts, AP/Honors Course enrollment, and Favorite Class). This parallels research that suggests a student's overall intelligence and familiarity with testing materials can be predictive of standardized test scores (Machudo, 2018). From the current study, as a student's GPA increases, so do their ACT scores. Students who take the ACT more times have higher ACT scores. Those participants who enrolled in AP/Honors courses scored higher than those who did not. Lastly, students whose favorite subject was one of the tested sections scored higher than those who preferred class was not on the ACT. Across the five regression analyses, academic variables were overwhelmingly predictive of the DVs with the strongest magnitudes. This was accentuated with GPA being the strongest predictor of the outcome variables in all five regression models.

The results of this study diverge in some respects from prior literature with regards to the most influential predictors in this study being those in which the participant has some influence over. Previous research has indicated that hereditary demographic variables provided the most influence on high-stakes standardized test results, in some cases explaining the majority of the variation in the outcomes (Edwards, 2006; Maylone, 2002; Rumberger & Palardy, 2005). However, in the current study, the results suggest that as a student moves away from more "inherited" demographic variables into more "nurtured" academic variables, the IVs' (i.e., not the main TEOP variables) influence over ACT scores increases. Specifically, the academically-related variables (i.e., GPA, ACT Attempts, AP/Honors Course enrollment, and Favorite Class) are under some level of control by the student. That is, the student can work towards improving his/her GPA through a variety of methods including extra studying, working with tutors, and asking the teacher for additional clarification. Similarly, a high school student (depending on various resources) can choose to take the ACT an unlimited number of times and decide to enroll in AP or Honors courses at their own discretion. Finally, a student's favorite class is strictly his/her individual preference. Thus, a student can "take control" of their own academic success by focusing on doing well in their classes, taking the ACT multiple times, and challenging themselves with more rigorous classes (i.e., whenever financially or practically possible).

The strong impact of Academic variables on ACT scores implies that a student's test results are not predetermined by the environment in which he/she was raised. While the IVs from the Demographic/Behavioral block were significant predictors of the DV, the stronger influence of the Academic IVs suggest that a student can overcome obstacles through hard work and good choices. This is supported by multiple studies that have evidenced AP students tend to earn higher standardized test scores than their non-AP peers (Ewing, Camara, & Mislsap, 2006; Matten, Shaw, & Xiong, 2009; McPhilip & Rawls, 2013). This study's outcomes support the belief that individual students can overcome more static, demographic variable "obstacles," and have the potential to otherwise influence their standardized test success and "alter" their future trajectory.

While student autonomy allows a strong-willed individual to overcome impediments, some academic variables are not completely under a high school student's control. GPA, for example, can be influenced by grade inflation and different grading scales across schools (National Center for Educational Statistics, 2015). Socio-economic status provides wealthier students with options to improve their scores, which their less well-off peers lack. Affluent families, for instance, can hire tutors to help with test preparation, as well as coaches to help the student write with college essays. This results in these students, with a combined family income exceeding $80,000 per year, outscoring their less wealthy peers (Jaschik, 2019). Additionally, the $50.50 fee to take the ACT presents an obstacle to many students and limits their opportunity to take it multiple times (ACT, 2017).

The TEOP was not a significant predictor of ACT outcomes. This could be due to test takers not distinguishing between standardized testing conditions and classroom testing conditions. The wording of the TEOP questions may have invited this overlap as the items begin with the words, "When taking a test..." It is reasonable to surmise that a test taker used their more common frame of reference of an in-class test, despite the IRB-approved informed consent form specifically indicating the study was guided by preferences for standardized testing environment. As validity is contextual (Kane, 2001), the participants' perception of what is being asked in the TEOP items is not what was initially conceptualized when the measure was developed, but rather an assessment of more general classroom assessments.

GPA, unlike standardized tests, is comprised of more than just one day's performance. GPA is an integrated measure of knowledge, intelligence, effort, ability to follow directions, and a limitless list of other IVs. There are numerous studies that suggest that high school GPA has a predictive relationship with college success (Geiser & Santelics, 2007; Noble & Sawyer, 2004). This significant relationship between the TEOP variables and GPA has several implications for various stakeholders such as improved study practices, the selection of peers for group test preparation, guided classroom instruction, and scheduling. All stakeholders can benefit from this finding by encouraging students' academic achievement and focusing on GPA in order to facilitate better outcomes on standardized test examinations. As the TEOP was not found to have a predictive relationship with high-stakes standardized test results (i.e., ACT), the measure can be used in isolation to better understand individual students' preferences. These abovementioned general implications will be detailed more specifically in the following paragraphs from the proximal (to the student) to distal levels.

Proximal

At the proximal level (i.e., the closest in proximity to the student), findings from this study may directly impact students and the parents/guardians of these students. Although the TEOP was uncorrelated with high-stakes standardized test outcomes (i.e., ACT), it can be used as a measure for students to be more introspective (i.e., a self-discovery tool) to potentially improve study habits (i.e., resulting in better academic performance such as during a classroom assessment). Students and parents/guardians can be made aware of the relationship between the environments they study under with classroom test conditions and the impact on GPA. Individual students can be informed that if they have a higher preference for "Sound," this could negatively impact their GPA, as they are most likely studying/preparing under distracting environmental settings. The human brain is incapable of multitasking and the process of studying while also listening to music or other sounds may distract the student from the primary goal of learning the material (Kirschner & Van Merrienboer, 2013). These students who enjoy auditory stimulation may be studying with music, or in crowded coffee shops, as one example. These conditions may be disrupting to the student and hinder their ability to concentrate.

This is unlike the general environments under which most assessments classroom or high stakes are given. These learners can be made aware of this link and advised of more appropriate study strategies.

Students can also form study groups that are based on their specific “Action” and “Sound” preferences. As compared to preparing alone, group studying can improve comprehension, help clarify confusing material, and motivate the learner (Oxford Learning, 2018). Students with low “Sound” predilections might consider seeking out similar students to learn with. On the other hand, those with higher “Sound” preferences could contemplate working with peers dissimilar to them. Thus they will limit the temptation to be distracted by outside noises.

Within the student’s home, the family can recognize the distraction of sound on the learner and provide silent time for studying, or to make a quiet area of the house available. These implications could be particularly useful to parents/guardians with limited experience with regards to higher education. The parents/guardians of students who will be the first generation to go to college or were themselves unsuccessful academically in school may not have the personal knowledge or resources to assist their children with high school subject matter.

Self-awareness garnered from the TEOP could be a useful tool for parents/guardians that home school their children. Researchers estimate that there are 1.69 million children being home schooled in the United States (Ray, 2018). This number represents 3.3% of the children between the ages of five and 17. As parents/guardians who home school their children are often not trained educators, the TEOP results can help them guide instruction and design assessments (National Center for Educational Statistics, 2016). The parent/guardian acting in the role of a teacher can use the “Action” or “Sound” preferences and tailor assessment strategies to the individual needs of their child. They now have evidence to support turning off the music or sounds. These results can give them the confidence to limit the auditory distractions.

Results from this study could have implications for students and parents/guardians enrolled in the emerging area of online schooling. In 2013-14 there were 2.7 million K-12 students enrolled in digital learning. This is an increase of 80% from the 2009-10 school year (Connections Academy, 2019). The projections are that the number of students enrolled in online schools will continue to increase. Digital learning allows for students to take class anywhere with an internet connection. The very nature of online schools gives the learner a lot of autonomy over the educational setting. The TEOP results can help shape their understanding of what conditions are most effective to optimize academic performance. Similar to students who are home schooled those enrolled in online schools or classes can use TEOP results to optimize their learning and assessment environments.

At the proximal level, the evidence from this study and other research (e.g., Fisher, Godwin, & Seltman, 2014; Gazzaley & Rosen, 2016; Zhang, Miller, Cleveland, & Cortina, 2018) advocates that auditory and physical distractions that occur during test preparation should be kept to a minimum if not completely eliminated while studying. The individual learner would be wise to prepare in a disruption-free environment and their families should strive to limit distractions around the student.

Distal

At the distal level (i.e., the farthest away in distance from the student), there are various groups of individuals with the potential to be impacted by the results from this study such as classroom teachers, high school building level administration, school district level administration, and online schools. This study's results suggest that distractions may negatively impact studying, which is predictive of classroom test performance (i.e., as one major component of GPA), and has useful implications for those who work with children. The results can be extrapolated for assessment and even scheduling to enhance the learning of high school students.

As discussed in the literature review, teachers are encouraged and expected to differentiate instruction to meet the individual needs of the learner. Differentiated instruction allows all students to access the same academic standards while providing educational strategies that are individualized to students' unique needs (Subban, 2006). Knowing that students have "Action" and "Sound" preferences can help teachers plan their lessons and assessments. As an example, when assigning groups for cooperative learning activities, the teacher can choose to pair students with comparable TEOP preferences. For decades, educational psychologists have proposed that student learning is improved when schoolwork is done collectively in classrooms as compared to alone at their desk (Schoenherr, 2006). These students are likely compatible and can work well together. This would allow the students to complement each other's strengths and weaknesses. Overall, future lessons and instructions could be planned with strategies designed to help the individual learner.

While a student's special education classification was not an IV of this study, there are still implications for special education teachers. Special education students are put on an Individual Education Plan (IEP) that is constructed by a team of teachers, the child's parents/guardians, and special education directors, among others. The IEP outlines objectives and benchmarks specific to a student who has a disability and/or requires accommodations (Baumel, 2016). Examples of IEP accommodations include but are not limited to extended time for assessments, having exams read to the student, and taking tests in settings outside of the classroom. Academic IEP goals are comprised of smaller objectives that the student can reasonably be expected to complete each school year (Baumel, 2016). The TEOP is a brief, nine-item measure that could provide insight into the special education student's testing style. Thus, the results can be used to garner a better understanding of the special education student's assessment "style," and define the strategies to help him/her meet the IEP goals.

Guidance counselors at many high schools serve as the testing coordinator and guide the students through the college application process (Brown, 1999). Although the TEOP has not been subjected to the rigors required for validating diagnostic testing like some scales used in medicine (i.e., the Mini-Mental State Examination [MMSE]; Pangman, Sloan, & Guse, 2000), guidance counselors can use the TEOP in a diagnostic manner. For example, guidance counselors can administer the TEOP to students at the beginning of the school year and intermittently throughout the academic calendar to shape educational strategies for individual students. School guidance counselors can give the TEOP to students who may be underperforming with regards to GPA and/or standardized test scores. Additionally, the TEOP results are in an easily "digestible" metric for guidance counselors and students to understand. This could facilitate a discussion on the "Action" and "Sound" preferences and if or how they guide the students' study habits. Overall, the guidance counselor can use this insight to educate an underperforming student on best practices for academic success.

As mentioned in the proximal level implications, previous research (Fisher, Godwin, & Seltman, 2014; Gazzaley & Rosen, 2016; Zhang, Miller, Cleveland, & Cortina, 2018) that suggests limiting distractions while learning is beneficial for students supports the results in this study. Educators both in the classroom and admnistrators should emphasize minimizing disruptions and provide stimulus-free opportunities in students' academic preparation.

Tracking is the process of assigning students into classroom groups by ability (Gamoran, 1992; Rubin, 2008). The effectiveness of tracking is hotly debated in academic circles. Proponents argue that it allows teachers to design lessons specific to the students' ability and allows for greater growth amongst high-ability students (Rogers, 1991). Opponents of tracking suggest it results in racial and social discrimination and lower engagement in the low-ability tracks (Hyland, 2008).One strategy for using the TEOP to build a schedule would be to pair students with teachers, whose teaching style matches the student's "Action and Sound" preferences, essentially tracking students by "Action" and "Sound" preferences. This partnering could result in increased engagement and motivation for the students. Logically, it follows that students who are more interested in their classes and put in more effort towards them will be more successful than disconnected classmates.

A related example would be that school districts can use TEOP scores to plan their course rosters. There may be two directions the school could choose here, and additional research would be necessary to indicate the best option. One choice would be to assign course rosters based on similar TEOP preferences. Research indicates that girls in single-gender schools benefit academically, particularly in the areas of math and science performance. Girls also report that there are fewer distractions in single-gender classrooms (Smyth, 2010). The teachers of these courses can then adapt instruction and assessment to the group inclination. A second choice would be to intentionally balance the TEOP preferences in the individual courses so as to not inadvertently have any individual class that leans too heavily "Action" or "Sound" in preferences.

Classroom diversity directly influences student learning outcomes (Queens University of Charlotte, 2019). Diverse environments encourage more robust classroom discussions, improved critical thinking, enhanced problem solving, and higher academic achievement (Henson & Eller, 2012; Siegel-Hawley, 2012).Scheduling classes with similar "Action" and "Sound" preferences in the same classes might benefit both groups as it could limit disruptions and permit teachers to design instruction around the group's preferences. Scheduling classes with diverse "Action" and "Sound" preferences could facilitate small group learning comprised of learners with divergent academic talents. It should be noted that the school administration implications from this section are most likely only feasible in larger schools. A small high school has limited options due to fewer students and teachers.

The main hypothesis in this study proposed that a student's individual "Action" and "Sound" preferences impact standardized test (i.e., ACT) results. This hypothesis was, in retrospect, too simplistic. There are potentially a variety of other variables involved in how individual preferences impact standardized test outcomes. The nonsignificant relationship between the TEOP and ACT results supports existing research that suggests there are multiple factors that influence learning including verbal and visual processing, as an illustration (Sternberg, 2006). For example, the theory of dual coding suggests that a person retains knowledge through verbal associations and visual imagery and that the merger of the two can enhance learning (Paivio, 1971). Dual coding can be impactful because it takes advantage of multiple mental processing channels.

The results from this study may support the concept of combining processes in education. As one example, studies have shown that adding dissection (i.e., a physical action) to medical school courses can improve both short- and long-term retention on quizzes and exams (Rae, Cork, Karpinski, Farris, & Swartz, 2015). With this in mind, there are multiple strategies for educators to embed sounds and actions related to the lesson objectives into the instruction. For instance, those students with higher "Action" preferences when learning Shakesphere in an English Language Arts or Literature course may benefit from acting out the plays as opposed to simply reading and discussing them. Similarly, student learning French with greater "Sound" preferences may benefit from listening to French music while studying or during instruction to provide an additional way to reinforce the learned material. Through the nonsignifcant findings in this study, the results may support previous research that demonstrates the importance of pairing "Action" or "Sound" preferences related to the lesson objective or learning process as one mechanism to enhance knowledge retention.

Limitations and Future Directions

Two sections of limitations are presented below: (1) Methodological/Statistical and (2) Psychometric. These limitations are accompanied by suggestions for future research.

Methodological/Statistical

The voluntary response sample is a limitation for the study, as it is impossible to substantiate self-reported data. Unfortunately, this is often the nature of Internet-based survey research (Kline, 2013). Self-reported information can be challenging because subjects may forget details, exaggerate, or answer based on perceived social desirability (Northrup, 1996).

Due to the communities from which the study schools were located and participants were recruited and the disproportionately Caucasian composition (84.9%) of the Main Study sample, the results may not be broadly generalizable (Dimitrov, 2010). The moderately homogeneous sample was in contrast to the demographic profile of U.S. high school students provided by the National Center from Education Statistics (NCES). In the U.S., the high school student population is approximately 48.9% Caucasian (NCES, 2017). The study's sample consisted of 64.1% private school students, this is compared to approximately 10.0% nationwide (NCES, 2017). Future studies should attempt to recruit a sample that is more representative of the population.

The divergent academic course offerings and extra-curricular options available at the schools in the sample can impact students' outcomes. Evidence suggests that test prep courses and tutoring lead to improved standardized test results (Valero-Cuevas, Sklaar, & Peters, 2019). The five schools in the study have independent selection of classes including AP and Honors courses. Additionally, some of the schools offer ACT/SAT Prep courses, provide Response to Intervention (National Center for Learning Disabilities, 2006), and embedded academic supports and some did not. The different academic selections may offer advantages in test preparation and remediation that benefits students, compared to the pupils in schools without these options. The five schools in the sample also provide distinctive extra-curricular offerings. Schools that offer programs like Mathletes, Creative Writing Clubs, and Academic Challenge teams may have provided tools to help students achieve academic and testing success. Future research might consider comparing schools with equivalent academic and elective offerings.

The study was limited by having no controls for teacher effectiveness. Research on standardized tests results submits that certain teachers are more effective than others and that kids taught by these more impactful teachers perform better on aptitude and achievement tests. There are countless variables that influence student achievement but of school related factors teacher quality is the most influential (Engberg, 2012). The subjects Math and Reading in particular are most influenced by teacher quality (Rand

Corporation, 2012). As both Math and Reading are tested subjects on the ACT, the abilty of the teachers within the individual schools, particualrly in the ACT subject areas (i.e., Math, Science, English, and Reading), could effect the varaition within the DVs. Future research might consider adding controls for teacher quality to the study.

The study only included proxies for socio-economic status (i.e., Mother's Level of Education and Father's Level of Education), and there were no additional IVs representative of home-life factors. Research indicates that compared to school factors individual and family inluences can be four to eight times more impactful on student achievement (Engberg, 2012). Future studies might consider including family and home-life variables to control for their possible influence. Similar to the above, is that this study did not have a Special Education IV. Special Education is designed to tailor student with individual needs on a continuum from learning disabilities to giftedness (McFarland, 2018). In order to receive special education services a student must be placed on an Individual Education Plan (IEP). The most recent data suggests that 13% of US students are receiving Special Education services (McFarland, 2018). Future studies should consider adding a Special Education predictor variable to the demographic items to capture the data of this group of students.

The number of variables that impact academic success and the difficulty in isolating them can be detrimental to using experimental research in educational settings. Additionally, experimental research in education presents ethical, cost, Hawthorne Effect (i.e., people have a tendency to alter behavior when being watched), and external validity issues (Schanzenbach, 2012) that can all limit educational research opportunities. Future

TEOP research, however, might consider conducting experimental studies in an attempt to look for a causal relationship between putting a student in their preferred conditions and test results. This can be done through random assignment to either a control or treatment group, followed by both groups taking a benchmarking measure, followed by a control group being given a released practice exam that mimics the ACT or an in class exam of interest under the traditional test conditions and an experimental group being given the same exam with accommodations for their preferred TEOP environments. The results can then be used to determine if aligning ones' test conditions with their preferences has a relationship with their results.

The IV GPA was the most influential variable in all five regression analyses. A disproportionate amount of variance in the respective DVs was explained by this one variable. Research on the relationship between GPA and ACT results indicates that high school GPA is positively predictive of ACT scores, specifically higher GPAs are predictive of increased ACT scores (Sawyer, 2010)**.** While it is inadvisable to eliminate a variable that measures overall academic success (like GPA) from research involving standardized test performance, the magnitude of influence explained by GPA presents a limitation. This presents a "Catch 22" situation that future investigations must consider based on the specific research goals.

The Main Analysis had 16 IVs to measure the TEOP's predictive relationship with ACT scores and the post hoc analysis used 13 IVs to predict the relationship between TEOP and GPA. Each additional predictor variable adds complexity to the research. One strategy to mitigate the complexity is to reduce the number of IVs, limiting the number of variables to only those critical to the research (Goggin, 1986). In order to simplify the study, additional research might consider identifying the IVs that are essential to the model and eliminating the others.

To simplify interpretation, 11 of the IVs in the five Hierarchical Multiple Linear Regressions with ACT Comprehensive and the individual subjects (i.e., Math, Science, English, and Reading) as the DV and nine of the IVs in the post hoc analysis that used GPA as the outcome were dichotomous. The use of dichotomous variables may limit the predictive specificity of the study due to treating every participant in one of two groups as the same (Royston, Altman, & Sauerbrei, 2006). For example, in the current study all students who exercised were included in the same group, regardless if they exercised one day a week or all seven. Furthermore, in the current study, the two variables measuring parents' education level were dichotomized to bachelor's degree or more and less than a bachelor's degree. Thus, a student whose father has a bachelor's degree in a low-earning potential field like Philosophy is treated the same as a participant whose father has a degree in Petroleum Engineering and high earning prospects (College Choice, 2019). The different levels of schooling within the same category have the possibility to place the participants in different socioeconomic classes despite their being measured equally in the model. The dichotomizing of ordinal and continuous variables might have impacted the magnitudes of the IVs on the respective DV. Future studies might want to consider using ordinal or continuous predictor variables to increase the authority of the study.

The possibility of grade inflation is another study limitation. Grade inflation is the tendency of some teachers to give students higher scores on tests, assessments, and report cards than the students deserve (Arenson, 2004). Evidence suggests that grade inflation is prevalent in U.S. High Schools, as demonstrated by larger number of students receiving top grades yet not meeting performance benchmarks on standardized assessments. Related to this, two-thirds of high school graduates are not academically prepared for college (Gershenson, 2018). Grade inflation can occur with individual teachers and/or at the school level. Therefore, a participant's high school of attendance or the teachers assigned to them may impact their GPA over and above the traditional factors of intelligence, work ethic, ability to follow instructions, and subject matter knowledge. The likelihood of grade inflation artificially influencing GPA in this study's sample limits the conclusions drawn. As GPA had the strongest relationship with DVs in the five regression analyses and was the DV in the post hoc study, inconsistent grading practices may limit the true magnitude of the relationships examined in this study. Future studies may consider using benchmarking tests or other standardized assessments to represent academic success or intelligence.

Each of the five high schools in the study has unique course offerings. This includes a variety of AP and Honors course. Therefore, the students in the varying schools do not have the same opportunity to take AP or Honors courses. This limits the study in several ways. One example is that since the schools do not offer the same number of AP and Honors courses the students did not have equal opportunity to take those classes. The participants in some schools can consequently take a larger number of the more rigorous AP and Honors classes than the students from the schools that offer less. Additionally, because of the differences in schedule offerings, the students' high school of attendance may not be able to enroll in AP or Honors courses in their preferred subjects, while a comparable student in a different school could. For example, a student with strong math skills may not attend a high school that offers AP Calculus and therefore cannot take the course, while a similar student at a different school can enroll in the class. Furthermore, different schools have varying criteria for taking AP and honors courses. Some schools allow anyone to take the courses regardless of academic performance, others require a prerequisite class first, some necessitate teacher recommendations, and others impose a minimum GPA to be eligible for AP and Honors classes. Future studies should consider using high schools with comparable course offerings from which to procure the study's sample.

Forty-seven participants were dropped from the CFA sample due to incomplete IV data, before the Hierarchical Multiple Linear Regressions were conducted. While the CFA and regressions ultimately had similar demographic results, the unequal sample sizes between the CFA and regressions were a limitation. The unequal sample sizes can result in a loss of influence and potentially result in Type I error (Keith, 2006; Rusticus & Lovato, 2014). To facilitate maximum predictive authority, future research should attempt to use nearly equivalent sample sizes.

To further the statistical analyses presented in this study, future research might consider using the Hierarchical Multiple linear Regression subject matter outcome variables (i.e., Math, Science, English, and Reading) or the individual ACT Composite to develop a Path Analysis Model. Path Models use observed variables and concurrently test numerous regression analyses. Path Analysis allows for the comparison of a sample correlation matrix to a theoretical model (Schumaker & Lomax, 2016). Following a supported Path Model, further research may consider combining the observed variables into a latent variable (i.e., for the subject matter ACT tests), allowing for Structual Equation Modeling (SEM) to be conducted. The primary difference between Path Analysis and SEM is that Path Analysis assumes that all varaibles are free of measurement error. Comparatively, SEM uses latent variables to account for measurement error (Schumaker & Lomax, 2016). Ultimately, future Path Analysis or SEM research could examine the sequential pathway related to the individual HMLR models conducted in this study (e.g., TEOP → GPA → ACT).

Psychometric

The TEOP CFA was replicated in two separate sub-populations (i.e., Millennial College Students and Generation Z High School Students). This replication provides evidence that the TEOP is a valid and reliable measure of an individual's preferred testing environment. The primary psychometric limitation of this research is that after examining the results the participants appear to view the TEOP as either a measure of classroom test environments or are unable to distinguish the difference between classroom and standardized testing environments. This was demonstrated by the TEOP not being predictive of any of the five ACT outcomes (i.e., Comprehensive, Math, Science, English, and Reading) but did have a significant impact on GPA when controlling for Demographic/Behavioral and Study/Testing variables. The language of the TEOP even states "When testing, I prefer...", not distinguishing between classroom tests and standardized assessments. It is possible that when people contemplate testing preferences their thoughts are of the more common in class variety as opposed to standardized assessments. Because of this future research might consider refining the TEOP usage to measure student preferences for in class assessments that occur daily in schools across the United States. That is the TEOP is probably more suited to measure environmental preferences of low stakes tests than standardized examinations to determine college aptitude.

This study involved students from both private and public schools. There are numerous differences between private and public schools. For example, in private schools, the parents' choose to send the student to a specific school compared to public school where attendance is assigned by the local school district. Within the individual schools, there are differences in policies regarding cell phone use, gum chewing, and bringing food and beverages to class among others. The students' exposure to their own school rules could shape how they rate preferences on the TEOP. For example, a student who is allowed to bring water to class may begin to rely on that environmental preference and rate having a beverage during testing on the TEOP as strongly necessary. On the other hand, a student who is not allowed to bring drinks to class may not even consider it a possibility and rate a low preference due to a lack of exposure. Forthcoming research might consider drawing from populations with comparable school rules and procedures.

Self-report measures or scales that require individuals to apprise on facets of their own personality, emotions, cognitions, or behaviors, can be problematic (Kazdin, 2003). While there are practical benefits to using self-report measures, this method of assessment heightens respondents' social desirability. The possibility of bias and distortion subjects involving their motives, self-interest, or to "look good" is increased with a self-report measure. This limitation of self-reporting results that shines the subject in the best light is referred to as the halo effect (Wiersma & Jurs, 2009). The halo effect is connected to the provocativeness of the measure. The constructs of interest measured in the TEOP are not divisive and students' preference would not be suggestive of status. Although the directions indicated that the results would not be shared with anyone specifically, and they would be reported collectively, participants may still have fretted over others' opinions. Future studies may consider deviating from self-report survey research methods and perhaps use direct observation or other measures.

Conclusion

The $4 billion a year testing industry in the United States has increased approximately threefold since the year 2000 and shows no signs of slowing down (Davis, 2016). The current study intended to enhance the limited body of theoretical and empirical literature examining standardized testing environmental conditions and the test takers conditional partialities. The goal of this study was to examine the test environment preferences amongst current high school students. This was a continuation of the Pilot Study that analyzed the test environment preferences for the current university population. The Main Study conducted a CFA to examine the psychometric properties of the TEOP in a new population and to provide further evidence to support the "Action" and "Sound" test environment affinities. Hierarchical Multiple Linear Regressions were conducted to provide evidence of the hypothesized relationships (i.e., Criterion-Related Validity – Concurrent) between the TEOP factors "Action" and "Sound" and aptitude outcomes (i.e., ACT scores). The results led to an additional post hoc Hierarchical Multiple Linear Regression using GPA as the DV.

The results of Main Study advanced the outcomes from the Pilot Study by providing additional validation support for the TEOP scores in a population of high school students. The study, therefore, provided validity and reliability evidence of the TEOP for two subpopulations (i.e., Millennial College Students and Generation Z High School Students). This suggests that current college and high school students have individual “Action” and “Sound “preferences of testing environments.

The findings also advise that while the TEOP did not have a significant predictive relationship with ACT Scores, it did have a significant relationship with a student’s GPA. This might ultimately prove to be more useful to stakeholders in improving student educational performance. The major testing companies are unlikely to modify their exam policies but those more proximal to the student have tremendous autonomy in designing classroom instruction and assessment. The TEOP requires no specialized equipment or training. It is an easy to use measure with a short administration period. People with interest in a high school student’s success either personally or professionally should interpret these results with caution, but can consider the use of the TEOP as a tool for in class assessment preparation.

APPENDICES

APPENDIX A

TEST ENVIRONMENT PREFERENCE FOR OPTIMAL PERFORMANCE

PILOT STUDY SURVEY

Appendix A

Test Environment Preference for Optimal Performance Pilot Study Survey

1. What is your age in years? (Round to the nearest whole number)

2. What is your gender?
    a. Male
    b. Female
    c. Other

3. I identify my race as…
    a. White
    b. Black/African American
    c. American Indian and Alaska Native
    d. Asian
    e. Hawaiian and Other Pacific Islander
    f. Hispanic/Latino
    g. Bi or Multiracial
    h. Other

4. What is your highest level of schooling?
    a. No high school diploma
    b. High School diploma
    c. Some college
    d. Bachelor's degree
    e. Master's degree
    f. Professional degree
    g. Doctorate degree

5. What is your employment status?
    a. Unemployed
    b. Work part time (less than 30 hours per week or 130 hours per month)
    c. Work full time (more than 30 hours per week or 130 hours per month)

6. What is your marital status?
    a. Single (never married)
    b. Married
    c. Separated
    d. Widowed

7. How many children do you have? If you have no children answer 0.

8. Including you how many people live in your household?

9. How do you describe your political views?
    a. Very Conservative
    b. Conservative
    c. Moderate
    d. Liberal
    e. Very Liberal

10. Did you take the ACT?
    a. Yes
    b. No

    If yes, what was your score?

    If yes to 10 above, approximately what year did you take the ACT?

11. If you are currently a college student what is your cumulative GPA?

12. My preferred time of day for taking a test is:
    a. 7:01 am-11:00 am
    b. 11:01 am- 3:00 pm
    c. 3:01 pm-7:00 pm
    d. 7:01 pm- 11:00 pm
    e. 11:01 pm-3:00 am
    f. 3:01 am-7:00 am

13. When taking a test I prefer silence:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

14. When taking a test I prefer background noise/environmental sounds such as, sound of the ocean, rain, birds chirping:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

15. When taking a test I prefer to listen to music:
    a. Never Me
    b. Rarely Me

b. Sometimes Me
c. Often Me
d. Always Me

16. When taking a test I prefer to have a television on in the background:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

17. When taking a test I prefer to have water/beverage available:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

18. When taking a test I prefer to have food available:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

19. When taking a test I prefer to chew gum or candy:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

20. When taking a test I prefer to have the option stand up:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

21. When taking a test I prefer to have the option to walk around:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me

c. Often Me
d. Always Me

22. When taking a test I prefer to take it on the computer:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

23. When taking a test I prefer to use paper and pencil:
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

APPENDIX B

QUALITATIVE INTERVIEW QUESTIONS (PILOT STUDY)

Appendix B

Qualitative Interview Questions (Pilot Study)

Introduction:

My name is Randy Rair, I am a graduate student at Kent State University, and I will be asking you some questions about your preferred testing environment. The interview should last approximately one hour. If you have any questions for me now or as we go on, please ask them. This interview is voluntary. You may decline to answer any or all of the questions and can quit at any time. After I have transcribed the interview, I may contact you again to ask a couple of follow up questions or to ask for clarification on your answers.

Student Interview Question Examples:

- Please describe yourself.
- Please describe the high school you went to.
- Did you take the ACT or SAT?
- If yes which one or both and when?
- What was your goal when sitting down to take the ACT/SAT?
- Can you describe your experience with the ACT/SAT?
- How would you describe the room you took the ACT/SAT in?
- How did you feel about the room?
- What did you like about the room?
- What did you not like during the test?
- What changes would you make to the room during the test?
- What was going on in the room when you took the ACT/SAT?
- How did you feel about what was going on in the room during the test?
- What was your experience with the test?
- What feelings or emotions did you have during the test?
- Outside of the material or content what is the most difficult part of taking a standardized test?
- While taking a test what do you do to help you to relax?
- While taking a test what do you do to help you to focus?
- While taking a test what do you do to help you perform better?
- Do you get nervous when taking a test?
- If you are nervous during a test what do you get nervous about?
- If you are nervous during a test what helps you relax when nervous?
- What things could distract you when taking a test?
- When taking a test what environmental things do you feel help you perform better?

- When you study describe the environmental setting?
- How does this compare with the ACT/SAT testing environment?
- When testing what makes you comfortable?
- When testing what makes you uncomfortable?
- What distracts you when taking a test?
- What helps you focus when taking a test?
- How do you feel about silence during a test?
- How would you feel about having background music being played during a test and how this might impact your performance?
- How would you feel having background noises being played during a test and how this might impact your performance?
- How would you feel having a television on in the background during a test and how this might impact your performance?
- Can you describe a time you were hungry during a test? How might this impact your performance?
- Can you describe a time you were thirsty during a test? How might this impact your performance?
- How do you feel having water available during a test might impact your performance?
- How would you feel about having food available to eat during a test might impact your performance?
- How would you feel about being allowed to chew gum during a test might impact your performance?
- How do you feel about being allowed to stand up during a test would impact your performance?
- How would you feel about being allowed to walk around during a test and would impact your performance?
- How would you set up the ideal conditions for you personally when taking a test so that you would perform your best?
- What would the ideal testing environment for you look, sound, and feel like?

APPENDIX C

TEST ENVIRONMENT FOR OPTIMAL PERFORMANCE SURVEY

Appendix C

Test Environment for Optimal Performance Survey

Please enter your ID Number:

1. What is your age in years? (Please round to the nearest whole number.)

2. What is your gender?
    a. Male
    b. Female
    c. Other

3. What is your racial or ethnic background?
    a. White
    b. Black/African American
    c. American Indian and Alaska Native
    d. Asian
    e. Hawaiian and Other Pacific Islander
    f. Hispanic/Latino
    g. Bi or Multiracial
    h. Other

4. What is your mother's highest level of schooling?
    a. No High School Diploma
    b. High School Diploma
    c. Some College
    d. Bachelor's Degree
    e. Master's Degree
    f. Professional Degree (e.g., Dentist, Lawyer)
    g. Doctoral Degree (e.g., Medical Doctor, PhD)
    h. Unknown

5. What is your father's highest level of schooling?
    a. No High School Diploma
    b. High School Diploma
    c. Some College
    d. Bachelor's Degree
    e. Master's Degree
    f. Professional Degree (e.g., Dentist, Lawyer)
    g. Doctoral Degree (e.g., Medical Doctor, PhD)
    h. Unknown
6. What is your employment status?
    a. Unemployed

b. Work part-time (i.e., less than 30 hours/week or 130 hours per month)
c. Work full time (i.e., more than 30 hours/week or 130 hours per month)

7. Including you, how many people live in your household?

8. How many times have you taken the ACT? (Do not include any practice ACT exams you may have taken. If you have not taken the ACT, please type “0” below.)

9. If you have previously taken the ACT, what was your score? (If you have taken the ACT more than once, please enter your best score.)

10. Have you taken any ACT prep courses?
a. Yes
b. No

11. Have you used a practice study guide to prepare for the ACT?
a. Yes
b. No

12. Do you participate in any extracurricular activities at your high school (e.g., athletics, speech and debate, Theater/Drama, other clubs, band, choir, etc.)?
a. Yes
b. No

13. If you answered “Yes” to the previous question, please list the extracurricular activities in which you participate.

14. Do you play a musical instrument?
a. Yes
b. No

15. Do you listen to music for enjoyment/relaxation?
a. Yes
b. No

16. If you answered “Yes” to the previous question, approximately how many hours per day do you listen to music for enjoyment/relaxation? (If you answered “No,” please select “N/A” below.)
a. Less Than 1 Hour/Day
b. More Than 1 Hour/Day
c. N/A

17. What is your current cumulative GPA? (Please enter the number from 0 to 4.0).

18. How many Advanced Placement courses have you taken? (Include courses you are currently taking. If you have not taken any Advanced Placement courses, please type "0" below.)

19. How many Honors courses have you taken? (Include courses you are currently taking. If you have not taken any Honors courses, please type "0" below.)

20. Which Math courses are you currently taking? (If you are not currently enrolled in a Math course, please type the name of the most recent Math course you have taken.)

21. Of the subjects listed below, which is your favorite?
    a. Math
    b. Science
    c. Social Studies
    d. English/Language Arts
    e. Art
    f. Physical Education

22. Do you study in silence?
    a. Yes
    b. No

23. Do you eat/drink/chew gum while studying? (Please select "Yes" if you do one or more.)
    a. Yes
    b. No

24. Do you exercise/workout on a regular basis?
    a. Yes
    b. No

25. If you answered yes to the previous question, on average, how many days per week do you exercise/workout? (If you answered "No," please select "N/A" below.)

    a. 1 Day/Week
    b. 2 Days/Week
    c. 3 Days/Week
    d. 4 Days/Week
    e. 5 Days/Week
    f. 6 Days/Week
    g. 7 Days/Week
    h. N/A

26. Do you play video/electronic games (i.e., including but not limited to on a video gaming system, tablet, laptop, cell phone, etc.)?
    a. Yes
    b. No

27. If you answered yes to the previous question, on average, how many days per week do you play video/electronic games (If you answered "No," please select "N/A" below.)?

    a. 1 Day/Week
    b. 2 Days/Week
    c. 3 Days/Week
    d. 4 Days/Week
    e. 5 Days/Week
    f. 6 Days/Week
    g. 7 Days/Week
    h. N/A

28. Do you plan to go to college/a university after high school?
    a. Yes
    b. No

29. Do you feel that you are a "good test taker?"
    a. Yes
    b. No

30. Do you have a difficult time staying focused when testing?
    a. Yes
    b. No

31. My preferred time of day for taking a test is…
    a. 7:01 am - 11:00 am
    b. 11:01 am - 3:00 pm
    c. 3:01 pm - 7:00 pm
    d. 7:01 pm - 11:00 pm
    e. 11:01 pm - 3:00 am
    f. 3:01 am - 7:00 am

32. When taking a test, I prefer silence.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

33. When taking a test, I prefer background noise/environmental sounds (i.e., sounds of the ocean, rain, birds chirping, etc.).
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me
34. When taking a test, I prefer to listen to music.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

35. When taking a test, I prefer to have a television on in the background.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

36. When taking a test, I prefer to have water/beverage available.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

37. When taking a test, I prefer to have food available.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

38. When taking a test, I prefer to chew gum or candy.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

39. When taking a test, I prefer to have the option stand up.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

40. When taking a test, I prefer to have the option to walk around.
    a. Never Me
    b. Rarely Me
    b. Sometimes Me
    c. Often Me
    d. Always Me

APPENDIX D

IRB APPROVAL PILOT STUDY QUANTITATIVE

Appendix D

IRB Approval Pilot Study Quantitative

Protocol #16-752 entitled "An Examination of a Performance Environment Questionnaire"

We have assigned your application the following IRB number: **16-752**. Please reference this number when corresponding with our office regarding your application.

The Kent State University Institutional Review Board has reviewed and approved your Application for Approval to Use Human Research Participants as Level I/Exempt from Annual review research. Your research project involves minimal risk to human subjects and meets the criteria for the following category of exemption under federal regulations:

- Exemption 2: Educational Tests, Surveys, Interviews, Public Behavior Observation

This application was approved on December 15, 2016.

******Submission of annual review reports is not required for Level 1/Exempt projects. We do NOT stamp Level I protocol consent documents.*

For compliance with:

- DHHS regulations for the protection of human subjects (Title 45 part 46), subparts A, B, C, D & E

If any modifications are made in research design, methodology, or procedures that increase the risks to subjects or includes activities that do not fall within the approved exemption category, those modifications must be submitted to and approved by the IRB before implementation. Please contact an IRB discipline specific reviewer or the Office of Research Compliance to discuss the changes and whether a new application must be submitted. Visit our website for modification forms.

Kent State University has a Federal Wide Assurance on file with the Office for Human Research Protections (OHRP); FWA Number 00001853.

To search for funding opportunities, please sign up for a free Pivot account athttp://pivot.cos.com/funding_main
If you have any questions or concerns, please contact us at Researchcompliance@kent.edu or by phone at 330-672-2704 or 330.672.8058.

Doug Delahanty | IRB Chair |330.672.2395 | ddelahan@kent.edu

Tricia Sloan | Coordinator |330.672.2181 | psloan1@kent.edu
Kevin McCreary | Assistant Director | 330.672.8058 | kmccrea1@kent.edu
Paulette Washko | Director |330.672.2704| pwashko@kent.edu

APPENDIX E

IRB APPROVAL PILOT STUDY QUANTITATIVE MODIFICATION

Appendix E

IRB Approval Pilot Study Quantitative Modification

RE: IRB # 16-752 entitled "An Examination of a Performance Environment Questionnaire"

Hello,

The Kent State University Institutional Review Board (IRB) has reviewed and approved your protocol amendment/change request. It is understood that the research is continuing with modifications including to offer extra credit for survey completion to students in Aviad Israeli's classes and include an alternative extra credit assignment (use revised consent forms). The modification to this protocol was approved on 2/21/17 for compliance with:

- DHHS regulations for the protection of human subjects (Title 45 part 46), subparts A, B, C, D & E

Federal regulations and Kent State University IRB policy requires that research be reviewed as full board (level III) or expedited (level II) at intervals appropriate to the degree of risk, but not less than once per year. Exempt applications do not require annual review.

HHS regulations and Kent State University Institutional Review Board guidelines require that any changes in research methodology, protocol design, or principal investigator have the prior approval of the IRB before implementation and continuation of the protocol. The IRB must also be informed of any adverse events associated with the study. The IRB further requests a final report at the conclusion of the study.

Kent State University has a Federal Wide Assurance on file with the Office for Human Research Protections (OHRP); FWA Number 00001853.

If you have any questions or concerns, please contact us at Researchcompliance@kent.edu or by phone at 330-672-2704 or 330.672.8058.

Thank you

Bethany Holland | Assistant |330.672.2384| bhollan4_stu@kent.edu
Tricia Sloan | Coordinator |330.672.2181 | psloan1@kent.edu
Kevin McCreary | Assistant Director | 330.672.8058 | kmccrea1@kent.edu
Paulette Washko | Director |330.672.2704| pwashko@kent.edu
Doug Delahanty | IRB Chair |330.672.2395 | ddelahan@kent.edu

APPENDIX F

IRB APPROVAL PILOT STUDY QUALITATIVE

Appendix F

IRB Approval Pilot Study Qualitative

RE: Protocol #17-279 entitled “Test Environment for Optimal Performance”

We have assigned your application the following IRB number: **17-279**. Please reference this number when corresponding with our office regarding your application.

The Kent State University Institutional Review Board has reviewed and approved your Application for Approval to Use Human Research Participants as Level I/Exempt from Annual review research. **This approval is good for 3 years from date of approval.** Your research project involves minimal risk to human subjects and meets the criteria for the following category of exemption under federal regulations:

- Exemption 2: Educational Tests, Surveys, Interviews, Public Behavior Observation

This application was approved on May 25, 2017.

****Submission of annual review reports is not required for Level 1/Exempt projects. We do NOT stamp Level I protocol consent documents.*

For compliance with:

- DHHS regulations for the protection of human subjects (Title 45 part 46), subparts A, B, C, D & E

If any modifications are made in research design, methodology, or procedures that increase the risks to subjects or includes activities that do not fall within the approved exemption category, those modifications must be submitted to and approved by the IRB before implementation.

Please contact an IRB discipline specific reviewer or the Office of Research Compliance to discuss the changes and whether a new application must be submitted. Visit our website for modification forms.

Kent State University has a Federal Wide Assurance on file with the Office for Human Research Protections (OHRP); FWA Number 00001853.

To search for funding opportunities, please sign up for a free Pivot account at http://pivot.cos.com/funding_main

If you have any questions or concerns, please contact us
at Researchcompliance@kent.edu or by phone at 330-672-2704 or 330.672.8058.

Doug Delahanty | IRB Chair |330.672.2395 | ddelahan@kent.edu
Tricia Sloan | Coordinator |330.672.2181 | psloan1@kent.edu
Kevin McCreary | Assistant Director | 330.672.8058 | kmccrea1@kent.edu
Paulette Washko | Director |330.672.2704| pwashko@kent.edu

APPENDIX G

IRB APPROVAL MAIN STUDY

Appendix G

IRB Approval Main Study

RE: IRB # 18-114 entitled "Test Environment for Optimal Performance in High School Students: Measure Development and the Relationship with Standardized Test Scores"

Hello,

I am pleased to inform you that the Kent State University Institutional Review Board reviewed and approved your Application for Approval to Use Human Research Participants as a Level II/Expedited, category 7 project. **Approval is effective for a twelve-month period:**

March 7, 2018 through March 6, 2019

For compliance with:

- DHHS regulations for the protection of human subjects (Title 45 part 46), subparts A, B, C, D & E

**If applicable, a copy of the IRB approved consent form is attached to this email. This "stamped" copy is the consent form that you must use for your research participants. It is important for you to also keep an unstamped text copy (i.e., Microsoft Word version) of your consent form for subsequent submissions.*

Federal regulations and Kent State University IRB policy require that research be reviewed at intervals appropriate to the degree of risk, but not less than once per year. The IRB has determined that this protocol requires an annual review and progress report. The IRB tries to send you annual review reminder notice by email as a courtesy. **However, please note that it is the responsibility of the principal investigator to be aware of the study expiration date and submit the required materials.** Please submit review materials (annual review form and copy of current consent form) one month prior to the expiration date. Visit our website for forms.

HHS regulations and Kent State University Institutional Review Board guidelines require that any changes in research methodology, protocol design, or principal investigator have the prior approval of the IRB before implementation and continuation of the protocol. The IRB must also be informed of any adverse events associated with the study. The IRB further requests a final report at the conclusion of the study.

Kent State University has a Federal Wide Assurance on file with the Office for Human Research Protections (OHRP); FWA Number 00001853.

To search for funding opportunities, please sign up for a free Pivot account at http://pivot.cos.com/funding_main

If you have any questions or concerns, please contact the Office of Research Compliance at Researchcompliance@kent.edu or 330-672-2704 or 330-672-8058.

Doug Delahanty | IRB Chair |330.672.2395 | ddelahan@kent.edu
Tricia Sloan | Coordinator |330.672.2181 | psloan1@kent.edu
Kevin McCreary | Assistant Director | 330.672.8058 | kmccrea1@kent.edu
Paulette Washko | Director |330.672.2704| pwashko@kent.edu

REFERENCES

REFERENCES

The College Board. (2017). *The SAT school day testing manual.* New York, NY: The College Board.

Abernethy, E. (1940). The effect of changed environemntal conditions upon the results of college examinations. *Journal of Psychology:Interdisciplinary and Applied 10*, 293-301.

Abrahams, F. (2015, July 29). *Understanding Generation Z learning styles in order to deliever quality learning experiences*. Retrieved from Precision Industries: http://www.precisionindustries.com.au/whats-hot-right-now-blog/understanding-generation-z-learning-styles-in-order-to-deliver-quality-learning-experiences

ACT. (2017, March 12). *ACT Test*. Retrieved from ACT Test Web Site: www.act.org

ADA. (1990). Americans with disabilities act. *Pub. L. No. 101-336, 104 Stat. 328* .

Alkharusi, H. (2013). Canonical correaltional models of students' perceptions of assessment tasks, motivational orientations, and learning strategies. *International Journal of Instruction Vol.6 (1)*, 21-38.

Allen, D., Ort, S., & Schimidt, J. (2009). Supporting classroom assessment practice: Lesson from a small high school. *Theory into Practice 48*, 72-80.

Allen, M. (2004). *Assessing academic programs in higher education.* San Francisco: Jossey-Bass.

Allen, M., & Yen, W. M. (2002). *Introduction to measurement theory.* Long Grove, IL: Waveland Press.

Alluri, V., Torben, P., Lund, E., Wallentin, M., Vuurst, P., Nandi, A., . . . Brattico, E. (2013). From Vivaldi to Beatles and back:Predicting lateralized brain responses to music. *NeuroImage*, 627-636.

Amazon.com. (2018, May 23). *Amazon.com*. Retrieved from Alexa: https://www.alexa.com/topsites

American Academy of Pediatrics. (2014). *School start times for adolescents.* Elk Grove, IL: American Acdemy of Pediatrics.

American Psychological Association. (2006, March 30). *American Psychological Association*. Retrieved from American Psychological Association: http://www.apa.org/research/action/multitask.aspx

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Education 84*, 261-271.

Anderson, K. (2007). Tips for teaching: Differentiating instruction to include all students. *Preventing School Failure 51 (3)*, 49-54.

Anderson, T., & Amemiya, Y. (1988). The asymptotic normal distribution of estimates in factor analysis under general covariance. *The Annals of Statistics, 16(2)*, 759-771.

Andrews, B., & Wilding, J. (2004). The realtion of depression and anxiety to life stress and achievement in students. *British Journal of Psychology 95*, 509-521.

Arenson, K. (2004, April 18). Is it grade inflation or are kids just smarter? *New York Times*.

Atkinson, T., Rosenfeld, B., Sit, L., Mendoza, T., Frusicone, M., Lavene, D., . . . Basch, E. (2012). Using confirmatory factor analysis to evaluate construct validity of brief pain inventory (BPI). *J Pain Symptom Manage*, 558-565.

Baddeley, A., Eysenck, W., & Anderson, M. (2009). *Memory.* New York, NY: Psychology Press.

Bandalos, D. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Mulitvariate Behavioral Research, 28 (3)*, 351-374.

Barnes and Noble College. (2015). *Getting to know Gen Z: Exploring middle and high schoolers expectations for higher education.* Elizabeth, NJ: Barnes and Noble College.

Barrett, M., Swan, A., Mamikonian, A., Ghajoyan, I., Kramarova, O., & Youmans, R. (2014). Technology in note taking and assessment: The effects of congruence on student performance. *International Journal of Instruction Vol 7 No. 1*, 1308-1470.

Baumel, J. (2016, December 19). *What is an IEP.* Retrieved from Great Schools: https://www.greatschools.org/gk/articles/what-is-an-iep/

Beall, G. (2017, November 6). *8 key differences between Gen Z and Millenials.* Retrieved from Huffington Post: https://www.huffingtonpost.com/george-beall/8-key-differences-between_b_12814200.html

Beastall, L. (2008). Enchanting a disenchanged child: Education using information and communication technology and elearning. *British Journal of Sociology of Education 27 (1)*, 97-110.

Beaton, C. (2017, January 27). *The millennial workforce: How multitasking is changing the workforce.* Retrieved from Forbes Magazine: https://www.forbes.com/sites/carolinebeaton/2017/01/27/the-millennial-workforce-how-multitasking-is-changing-our-brains/#1733ccb93605

Bentler, P., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin 88*, 588-606.

Bogdan, R., & Biklen, S. K. (1998). *Research in Education (3rd Edition).* Needham Heights, MA: Allyn and Bacon.

Booth, P., Taylor, B., & Edmonds, C. (2012). *Water supplementation improves visual attention and fine motor skills in school children.* London: National Hydration Council.

Brookhart, S., & Nitko, A. (2015). *Edcuational Assessments of Students.* Upper Paddle River, New Jersery: Pearson.

Brown, D. (1999). *Improving academic achievement: What school counselors can do.* Greensboro, NC: UNC Greensboro.

Brown, J. (2009). Choosing the right type of rotation in PCA and EFA. *Shiken: JALT Testing and Evaluation SIG Newsletter 13(3)*, 20-25.

Bryant, F., & Yarnold, P. (1995). Prinicpal-components analysis and confirmatory factor analysis. In L. Grimm, & P. Yarnold, *Reading and understanding multivariate statistics* (pp. 99-136). Washington, DC: American Pyschological Association.

Camstra, A., & Boomsma, A. (1992). Cross-validation in regression and covariance structure analyis. *Sociological Methods and Research, 21 (1)*, 89-115.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Charmaz, K. (2014). *Constructing grounded theory.* Los Angeles, CA: Sage.

Chazdon, S., Allen, R., Horntvendt, J., & Scheffert, D. (2013). *Methodological appendix: Steps before confirmatory factor analysis.* Minneapolis, MN: University of Minnesota.

Chingos, M. (2012). *Standarized in numbers: State spending on K-12 assessment system.* Washington, DC: Brookings Institute.

Chun, M., Golomb, J., & Turk-Browne, N. (2011). A taxonomy of external and internal attention. *Annual Review Psychology* , 73-101.

Clark, K. (2015). *Maintaining cognitive & psychometric integrity of testing instruments in translation.* Jersey City, NJ: Responsive Translation Services.

Clark, L., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. . *Psychological Assessment 7*, 309-319.

Claveria, K. (2018, August 28). *Vision Critical.* Retrieved from Vision Critical: https://www.visioncritical.com/generation-z-statistics/

College Choice. (2019, March 18). *College Choice*. Retrieved from College Choice: https://www.collegechoice.net/50-highest-paying-careers-college-graduates/

Comrey, A. J. (1988). Factor-analytic methods of scale development in personality and clinical pyschology. *Journal of Consulting and Clinical Psychology 56*, 754-761.

Comrey, A., & Lee, H. (1992). *A first course in factor analysis.* Hillsdale, NJ: Lawrence Earlbaum.

Connections Academy. (2019, March 9). *Coonections Academy*. Retrieved from Connections Academy: https://www.connectionsacademy.com/news/growth-of-k-12-online-education-infographic

Copper, R., Seppanen, S., & Gualiteri, W. (2012). *Millennial generation research review.* Washington, DC: United States Chamber of Commerce Foundation.

Corbalan, G., Kester, L., & Merrenboer, V. (2006). Towards personalized task selection model with shared instructional control. *Instructional Science 34*, 399-422.

Corbalan, G., Kester, L., & Van Merrienboer, J. (2008). Selecting leaning tasks:Effects of adaptation and shared control on learning effeciency and task involvement. *Contemporary Educational Psychology 33*, 733-756.

Corbalan, G., Kester, L., & Van Merrienboer, J. (2009). Combining shared control with variabilityover surface feature effects on transfer test performance and task involvement. *Computers in Human Beavior 25*, 290-298.

Cortes, K., Bricker, J., & Rohlfs, C. (2010). The role of specific subjects in education production functions: Evidence from morning classes in Chicago public high schools. *The BE Journal of Economic Analysis and Policy*.

Cortina, J. (1993). What is coefficient alpha? An examiniation of theory and applications. *Journal of Applied Psychology*, 98-104.

Costello, A., & Osborne, J. (2005). Best practices in Exploratory Factor Analysis:Four reccommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation,10 (7)*, 1-9.

Cox, J. (2017, Febuary 1). Google is now the world's most valuable brand. *The Independent*.

Cresswell, J. W., & Plano Clark, V. (2011). *Designing and conducting mixed method research (2nd ed.).* Thousand Oaks, CA: Sage.

Creswell, J. (2007). *Qualitative inquiry and research design: Choosing among five approaches (2nd. ed).* Thousand Oaks, CA: Sage.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory.* Mason, Ohio: Cengage Learning.

Cronbach, L. (1951). Coeffecient alpha and the internal structure of tests. *Psychometrika 16 (3)*, 297-334.

Cronbach, L., & Meehl, P. (1955). Construct validity in pyschological tests. *Psychological Bulletin*, 281-302.

Crotty, M. (1998). *The foundations of social research.* Thousand Oaks, CA: Sage.

Crowley, S., Van Reen, E., LeBourgeois, M., Acebo, C., Tarokh, L., & Seifer, R. (2014). A longitudinal assessment of sleep timing, circadian phase, and phase angle of entrainment across human adolescence. *PLos One*.

Cudeck, R., & Browne, M. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research 18*, 147-167.

D'Agostino, J., & Powers, S. (2009). Predicting teacher performance with test scores and grade point averages: A meta-analysis. *American Educational Research Journal 46(1)*, 146-182.

Dancer, D., & Kamvounias, P. (2005). Student involvement in assessment: A project designed to assess class participation fairly and reliably. *Assessment and Evaluation in Higher Education 30*, 445-454.

Davis, O. (2016). *No test left behind.* Retrieved from Talking Points Memo: https://talkingpointsmemo.com/features/privatization/four/

Dimcock, M. (2018). *Defining generations:Where Millenials end and post-Millenials begin?* Washington DC: Pew Research Center.

Dimitrov, D. (2010). *Quantitative Research in Education.* Oceanside, NY: Whittier Publications.

Dochy, F., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation 23*, 279-298.

Edmonds, C., & Buford, D. (2009). Should we drink more water?: The effects of drinking water on cognition in children. *Appetite 52 (3)*, 776-779.

Edmonds, C., & Jeffes, B. (2009). Does having a drink help you think? *Appetite 53 (3)*, 469-472.

Edwards, F. (2012). Early to Rise? The Effect of Daily Start Times on Academic Performance. *Economics of Education Review*, 970--983.

Edwards, H. (2015, June 29). *Prep Scholar*. Retrieved from Prep Scholar: https://blog.prepscholar.com/why-is-the-act-called-the-american-college-test

Edwards, N. (2006). The historical and social foundations of standardized testing: In search of balance between learning and evaluation. *Shiken and Shiken Research Journal: JALT Testing and Evaluation 10*, 8-16.

Engberg, J. (2012). *Teachers matter: Understanding teacher impact on student achievement.* Santa Monica, CA: Rand Corporation.

Ewing, M., Camara, W., & Mislsap, R. (2006). *The realtionship between PSAT/NMSQT scores and AP examiniation grades.* New York, NY: The College Board.

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use exploratory factor analysis in Psychological research. *Psychological Methods,4 (3)*, 272-299.

Ferinez, A., Hortovanyi, L., Szabo, R., & Tarody, D. (2010). *Changes in the way of work: Generation Z at the labour market.* Budapest: Corvinus.

Field, A. (2014). *Discovering statistics using IBM SPSS statistics 4th edition.* London: Sage Publications.

Fisher, A., Godwin, K., & Seltman, H. (2014). Visual entertainment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science 25 (7)*, 1362-1370.

Fowler, F. (2002). *Survey research methods.* Thousand Oaks, CA: Sage Press.

Frey, B., & Schimitt, V. (2007). Teacher classroom assessment practices. *Middle Grades Research Journal 5 (3)*, 107-117.

Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences 7 (3)*, 385-400.

Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership 50 (2)*, 11-17.

Gazzaley, A., & Rosen, L. (2016). *The distracted mind: Ancient brains in a high-tech world.* Cambridge, MA: The MIT Press.

Gee, J. (2003). *Good video games and good learning.* Madison, WI: University of Wisconsin.

Geiser, S., & Santelics, M. (2007). *Validity of high school grades beyond freshman year: High school record vs. standardized test as an indicator of four year college outcomes.* Berkely, CA: Berkely Education Publishers.

George, D., & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference 17.0 (10th ed).* Boston: Pearson.

George, S., Ahmed, S., Mammen, K., & John, G. (2011). Influence of music on the operation theatre staff. *Journal of Anaesthesilogy Clinical Pharmacology, July-Spetember 27(3)*, 354-357.

Gershenson, S. (2018). *Grade inflation in high school (2005-2016).* Washington, DC: Fordham Institute.

Gewertz, C. (2017, September 7). Students' scores inch up on the ACT exam. *Education Week*.

Gewertz, C. (2017, Febuary 15). Which states require student's to take the SAT or ACT? *Education Week*.

Gibbons, M. (2002). *The self-directed learninghandbook:Challenging adolescent students to excel.* San Francisco: Josey-Bass.

Gietz, C. (2008). *Relationship between student perception of their school environment and academic achievement.* Univeristy of Northern British Columbia.

Godden, D., & Baddeley, A. (1975). Context dependent memory in two natural environments: On land and underwater. *British Journal of Psychology 66*, 325-331.

Goggin, M. (1986). The "too few cases/too many variables" problem in implementation research. *The Western Political Quarterly 39(2)*, 328-347.

Goodson, I., Lankshear, M., & Mangan, C. (2002). *Cyberspaces/Social Spaces: Culture clash in computerized classroom.* London: Palgave McMillan.

Goodwin, D. (1969). Alcohol and recall:State-dependenteffects in man. *Science 163*, 1358-1360.

Gopinathan, P., & Sharma, V. (1988). Role of dehydration in heat stress-induced variations in mental performance. *Archives of Environmental Health 43*, 15-17.

Gorsuch, R. (1983). *Factor Analysis.* Hillsdale, NJ: Lawrence Earlbaum.

Graetz, K. (2006). The Psychology of Learning Environments. *Educause Review*, 60-75.

Grant, H., Bredahl, L., Clay, J., Groves, G., McDorman, T., & Dark, V. (1998). Context dependent memory for meaningful material:Information for students. *Applied Cognitive Psychology 12*, 617-623.

Gregorie, C. (2014). Why surgeons listen to music in the operating room and how it could help you. *Wellness*.

Gundlach, M. (2017, July 11). *Bright Hub Education.* Retrieved from Bright Hub Education: http://www.brighthubeducation.com/teaching-methods-tips/106939-history-of-differentiated-instruction/

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika 19*, 149-161.

Hansen, M., Janssen, I., Schiff, A., & Phyllis, C. (2005). The impact of school daily schedule on adolescent sleep. *Pediatrics*.

Hart, R., Casserly, M., Uzzell, R., Palacios, M., Corcoran, A., & Spurgeon, L. (2015). *Student testing in America's great city schools: An inventory and preliminary analysis.* Washington, DC: Council of Great City Schools.

Hatch, J. A. (2002). *Doing qualitative research in education settings.* Albany, NY: State University of New York.

Hatfield, R. (2017, July 7). *Dr. Hatfield's Website*. Retrieved from Dr. Hatfield's Website: http://www.dr-hatfield.com/

Hativa, N., & Bienbaum, M. (2000). Who prefers what? Disciplinary differences in students' preferred approaches to teaching and learning styles. *Research in Higher Education, 41(2)*, 209-236.

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY: Guilford Press.

Heick, T. (2018, July 29). How Google impacts the way students think. *Teach Thought*.

Henson, K., & Eller, B. (2012). *Educational Psychology 2nd. Edition.* Dubuque, IA: Kendall Hunt Publishing.

Hess, K. (2015, January 3). Why do we make students sit still in class? (C. Blatt-Gross, Interviewer)

Hill, K., & Wigfield, A. (1984). Test Anxiety: A Major Educational Problem and What Can be Done About it. *The Elementary School Journal* , 107-125.

Hiss, W., & Franks, V. (2014). *Defining promise:Optional standardized testing policies in American colleges and university admissions.* Arlington, Va: National Association of College Admission Counseling.

Howard, G. (2007, March). As diversity grows, so must we. *Educational Leadership*.

Howe, N., & Strauss, W. (2009). *Millennials rising:The next generation.* New York: Knopf Doubleday Publishing Group.

Hu, L., Bentler, P., & Kano, Y. (1992). Can test statistics in covariance analysis be trusted? *Psychological Bulletin 112(2)*, 351.

Hunt, M. R. (2009). Strengths and challenges in the use of interprative descriptions: Reflections arising from a study of the moral experience of health professionals in humanitarian work. *Qualitative health research 19(9)*, 1284-1292.

Hyland, N. (2008). Detracking in Social Studies: A path to a more democratic education? *Theory into Practice 45 (1)*, 64-71.

IBM. (2011). *IBM Knowledge Center*. Retrieved November 28, 2016, from IBM Website: https://www.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/idh_cpca.htm

Ivankova, N., Creswell, J., & Stick, S. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods,18 (1)*, 3-20.

Jacobsen, E. (2018, September 11). *Erik the Red*. Retrieved from Erik the Red: https://www.erikthered.com/tutor/sat-act-history.html

Jacobsen, W., & Forste, R. (May 2011). The wired generation: Academic and social outcomes of electronic media use among university students. *Cyberpsychology, Behavior, and Social Networking 14 (5)*, 275-280.

Jaconi, M. (2017). The on-demand economy is revolutionizing consumer beahvior. *Tech Insider*.

Jaschik, S. (2019, March 2019). *Inside Higher Ed.* Retrieved from Wealth and admissions: ttps://www.insidehighered.com/admissions/article/2019/03/18/look-many-legal-ways-wealthy-applicants-have-edge-admissions

Jennings, R., Nebs, R., & Brock, K. (1988). Memory retrieval in noise and psychophysiological response in the young and old. *Psychophysiology vol.25 no.6*, 633-644.

Johnson, R. (2008, July 21). *Tool targets faulty 'one size fits all computer interfaces'.* Retrieved from EE Times: https://www.eetimes.com/document.asp?doc_id=1309668

Johnston, C., Tyler, C., Stansberrry, S., Palcic, J., & Foreyt, J. (2009). Gum cheiwng effects academic performance in adolescents,. *ASN Scientific Sessions and Annual Meeting at Experimental Biology.* New Orleans: ASN Scientific.

Joreskog, K., & Sorbom, D. (1996). *LISREL 8: User's reference guide.* . Scientific International.

Kalkhurst, D. (2018, March 12). Engaging Gen Z students and learners. *Higher Education*.

Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement 38*, 319-342.

Kapusciniski, A., & Masters, K. (2010). The currentstatus of measures of spirituality: A critical review of scale development. *Pschology of religion and spirituality 2(4)*, 191-205.

Karpniski, A., Kirschner, P., Mellott, J., & Ochwo, P. (2013). An exploration of social networking site use, multitasking, and academic performance among United States and European university students. *Coumputers in Human Beahvior 29(3)*, 1182-1192.

Kayluga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to optimize the effeciency of e-learning. *Educational Technology, Reserearch and Development 53(3)*, 83-93.

Kayluga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to optimize the effeciency of e-learning. *Educational Technology, Reserearch and Development 53(3)*.

Kazdin, A. E. (2003). *Research in design in clinical psychology.* Boston, MA: Allyn & Bacon.

Keith, T. (2006). *Multiple regression and beyond.* Boston: Pearson.

Kelley, P., Lockley, S., Foster, R., & Kelley, J. (2015). Synchronizing education to adolsecent biology: Let teeens sleep, start school later. *Learning Media and Technology 40(2)*.

Kinsman, J. (2014, February 1). Technology makes the world smaller, but isn't that a shame. *The Independent*.

Kinzie, M., & Sullivan, H. (1989). Continuing motivation, leaner control, and CAI. *Educational technology, Research and Development 37(2)*, 5-14.

Kinzie, M., Sullivan, H., & Berdel, R. (1992). Motivational and achievement effects of learner control over content review with CAI. *Journal of Educationg Computer Research Vol. 8 Issue 1*, 101-114.

Kirschner, P., & Van Merrienboer, J. (2013). Do learners really kow best? Urban legends in education. *Educational Pyschologist 48(3)*, 169-183.

Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist discovery, problem based, experimental, and inquiry based teaching. *Educational Psychologist 41(2)*, 75-86.

Kise, J. (2007). *Differentiation through personality types: A framework for instruction, assessment, and classroom management.* Thousand Oaks, CA: Corwin.

Kise, J. (2011, Summer). *Let me learn my own way.* Retrieved from Educational Leadership: http://www.ascd.org/publications/educational-leadership/jun11/vol68/num09/Let-Me-Learn-My-Own-Way.aspx

Kline, R. (2013). Exploratory and confrimatory factor analysis. In *Psychology* (pp. 171-205). Concordia, CA: Concordia University.

Koepp, G., Snedden, B., Flynn, L., Huntsman, B., & Levine, J. (2012). Feasibility analysis of standing desks for 6th graders. *Childhood Obesity and Nutrition 4(2)*, 89-92.

Krecar, I., Kolega, M., & Kunac, S. (2014). Effects of drinking water on attention. *Procedia Social Science and Behavioral Sciences 159*, 577-583.

Laerd Statistics. (2017, November 8). *Lund Research Limited*. Retrieved from Laerd Statistics: https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

Lahey, G. (1976). Learner control of lesson strategy:A model for Plato IV system. *ERIC #ED*, 125-143.

Lambert, N. (2018, February 23). An inconvienient truth:Gen Z and the culture of instant gratification. *Sapho*.

Larocque, M. (2008). Assessing the perception of the environemtn in elementary classrooms:The link with Achievement. *Educational Psychology in Practice 24(4)*, 289-305.

Lawrence, G. (2009). *People types and tiger stripes (4th ed.).* Gainesville, FL: Center for Applications of Psychological Type.

Lawrence-Brown, D. (2004). Differentiated instruction:Inclusive strategies for standards-based learning that benefit the whole class. *American Secondary Education 32(3)*, 34-62.

Layton, L. (2015, October 24). Study says standardized testing is overwhelming nation's public schools. *Washington Post*.

Lehman, A. (2005). *For basic univariate and multivariate statistics:A step-by-step guide.* Cary, NC: SAS Press.

Lepp, A., Barkley, J., & Karpinski, A. (2014). The relationship between cell phone use, academic performance, anxiety, and satisfaction with life in college students. *Human Behavior 31*, 343-350.

Lichtenstein, S., & Solvic, P. (2006). *The construction of preference.* New York, NY: Cambridge University Press.

Lipman, A. (2010, July 5). *Comscore releases November 2009 U.S search engine rankings*. Retrieved from ComScore.com: https://www.comscore.com/Insights/Press-Releases/2009/12/comScore-Releases-November-2009-U.S.-Search-Engine-Rankings?cs_edgescape_cc=US

Loevinger, J. (1954). The attention paradox in test theory. *Psychological Bulletin 51*, 493-504.

Lomax, R., & Hahs-Vaughn, D. (2012). *An introduction to statistical concepts.* Taylor & Francis Group: New York, NY.

Machudo, M. (2018, May 22). *Factors that affect student test scores*. Retrieved from Owlcation: https://owlcation.com/academia/Factors-That-Affect-Students-Test-Scores

Mangles, C. (2018, January 30). *Search Engine Statistics*. Retrieved from Smart Insights: https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/

Mark, G. (2008). *The cost of interrupted wor:More speed and stress.* Irvine, CA: University of California at Irvine.

Marklein, M. (2007, March 18). All four-year U.S. colleges now accept ACT. *USA Today*.

Marshall, C., & Rossamn, G. B. (1995). *Designing Qualitaitve Research.* Thousand Oaks, CA: Sage.

Mastin, L. (2018, December 4). *The Human Memory.* Retrieved from Short Term (Working) Memory: http://www.human-memory.net/types_short.html

Matten, K., Shaw, E., & Xiong, X. (2009). *The realtionship between AP exam performance and college outcomes.* 2009: The College Board.

Maylone, N. (2002). *The realtionship of socioeconomic factors and district scores on Michigan education assessment program tests: An analysis.* Ypsilanti, MI: Eastern Michigan Univeristy .

McCrindle, M. (2016). *Generation Z*. Retrieved from Generation Z: gnerationz.com.au

McDonald, A. (2001). The prevalence and effects of test anxiety in school children. *Educational Pyschology 21 (1)*, 89-101.

McFarland, J. (2018). *Children and youth with disabilities.* Washington, DC: National Center for Education Statistics.

McGlynn, A. (2005). Teaching millennials our newest cultural cohort. *Education Digest, 71(4)*, 12-16.

McGrindle Research. (2010). *Word Up*. Retrieved from Generation Z Web Site: http://www.generationz.com.au/pub/wordup1.pdf

McPhilip, M., & Rawls, A. (2013). A closer examiniation of the academic benefit of AP. *Journal of Educational Research 106*.

Meador, D. (2017, March 18). *High stakes testing: Overtesting in America's public schools.* Retrieved from Thought Co : https://www.thoughtco.com/high-stakes-testing-overtesting-in-americas-public-schools-3194591

Mehta, R., Shortz, A., & Benden, M. (2015). Standing up for learning: A pilot investigation on neurocognitive benefits of standing school desks. *International Journal of Environmental Research and Public Health 13 (2)*.

Merriam Webster. (2018, September 6). *Merriam Webster*. Retrieved from Merriam Webster: https://www.merriam-webster.com/dictionary/assessment

Merriam, S. (2002). *Qualitative research in practice.* San Francisco: Jossey-Boss.

Messick, S. (1995). Validity of psychological assessment:Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Microsoft Corporation. (2015). *Microsoft Corporation.* Retrieved from Microsoft Corporation Website: https://www.scribd.com/document/265348695/Microsoft-Attention-Spans-Research-Report

Montilla, J. (2004). *The construct validation of an instrument based on students' university choice and their perceptions of professor effectiveness and academic reputation at the Univeristy of Los Andes.* Los Andes, Chile: Retrieved from http://scholarcommons.usf.edu/etd/1166.

Moore, W. (1989). The learning evironment preference: Exploring construct validity of an objective measure of the Perry scheme of intellectual development. *Journal of College Student Development, 30*, 504-514.

Morgado, F., Meireles, J., Neves, C., Amaral, A., & Ferriera, M. (2017). Scale development: Ten limitations and reccommendations to improve future research practices. *Psicologia: Reflexao e Critica, 30(3)*, 1-20.

Morgan, K., Johnson, A., & Miles, C. (2013). Chewing gum moderates the vigilance detriment. *British Journal of Psychology* .

Mueller, C., & Kim, J. (1978). *Introduction to factor analysis:What it is and how to do it.* Beverly Hills, CA: Sage.

Myers, J., & Well, A. (2003). *Research design and statistical analysis (2nd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

National Center for Education Statistics. (2017, April 16). *National Center for Education Statistics*. Retrieved from National Center for Education Staitstics Web Site: https://nces.ed.gov/programs/digest/d15/tables/dt15_306.20.asp?current=yes

National Center for Educational Statistics. (2015). *How is grade average calculated?* Washington, DC: United States Department of Education.

National Center for Educational Statistics. (2016). *Homeschooling rate by parental education.* Washington, DC: United States Department of Education.

National Center for Learning Disabilities. (2006). *A parent's guide to response to intervention.* New York, NY: National Center for Learning Disabilities. Retrieved from RTI Action Network.

National Education Association. (2017). *Indicators of future success: GPA and noncognitive skills.* Washington, DC: National Education Association.

National Science Foundation. (2015). *Doctorate recipients from U.S. universities.* Washington, DC: National Center for Science and Engineers Statistics Directorate.

Neal, B. (2018, March 11). *9 differences between older and younger millenials.* Retrieved from Bustle: https://www.bustle.com/p/9-differences-between-older-millennials-younger-millennials-67779?utm_source=facebook&utm_medium=pro&utm_campaign=fbpro67779

Neil, J. (2017). *Survey research design in Psychology.* Canberra, Australia: Creative Commons.

Nielsen Company. (2012, March 29). *Nielsen.com*. Retrieved from Nielsen Company: http://www.nielsen.com/us/en/insights/news/2012/smartphones-account-for-half-of-all-mobile-phones-dominate-new-phone-purchases-in-the-us.html

Noble, J., & Sawyer, R. (2004). *Predicting different levels of academic success in college using high school GPA and ACT composite score.* Iowa City, IA: ACT Research.

Northrup, D. (1996). The problem of self-report in survey research. *The Institute for Social Research 11(3)*.

Nunley, K. (2004). *Layered Curriculum (2nd ed.).* Amherst, NH: Brains.org.

Nunley, K. (2006). *Differentiating the high school classroom:Solution strategies for 18 common obstacles.* Thousand Oaks, CA: Corbin.

Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw Hill.

Nunnally, J., & Bernstein, I. H. (1994). *Psychomteric theory (3rd ed.).* New York, NY: McGraw-Hill.

Oblinger, D. (2008). Growing up with Google: What it means to education. *Net Generation Emerging Technologies, 3*, 1-21.

Olson, A., & Sabers, D. (2008). *21st Century education: A reference handbook.* Thousand Oaks, CA: Sage Publications.

Ophir, E., Nass, C., & Wagner, A. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences of the United States of America Vol 106 no 37*, 15583-15587.

Oppezzo, M., & Schwartz, D. (2014). Give your ideas some legs: The positive effect of walking on creative thinking. *Journal of Experimental Psychology, 40 (4)*, 1142-1152.

Osborne, J., & Blanchard, M. (2010). Random responding from participants is a threat to validity of social science research results. *Frontiers in Pyschology, 1*, 220-225.

Oxford Learning. (2018, 11 26). *Studying in groups vs. studying alone. Which is better?* Retrieved from Oxford Learning: https://www.oxfordlearning.com/studying-alone-vs-studying-in-a-group/

Paivio, A. (1971). *Imagery and verbal process.* New York, NY: Holt, Rinehart, and Winston.

Pal, E. (2013). *The generation z.* Pecs, Hungary: The University of Pecs.

Pangman, V., Sloan, J., & Guse, L. (2000). An examination of pyschometric properties of the Mini-Mental Status Examination and the Standardized Mini-Mental Status Examination: Implications for C. *Applied Nursing Research 13 (4)*, 209-2013.

Parker, A., & Gellatly, A. (1997). Moveable cues: A practical method for reducing context-dependent forgetting. *Applied Cognitive Psychology 11*, 163-173.

Pashler, H. (1994). Dual task interference in simple tasks: Data and theory. *Psychological Bulletin 116 (2)*, 220-244.

Passion, J. (2017, September 28). Music, rhythm, and the brain. *Brain World Magazine*.

Pawson, C., Doherty, S., Martin, L., Soares, R., & Edmonds, C. (2012). Students who bring water into exams may improve their grade by keeping hydrated. *British Psychological Society Annual Conference.* London : British Psychological Society.

Pedhazur, E., & Schmelkin, L. (1991). *Measure, design, and analysis:An integrated approach.* Hillsdale, NJ: Lawrence Earlbaum.

Pelham, W., Waschbusch, D., Hoza, B., Gnagy, E., Greiner, A., Sams, S., . . . Carter, R. (2011). Music and video distractors for boys with ADHD in the classroom:comparison with controls, individual differences, and medication effects. *Abnormal Child Psychology 39 (8)*, 1085-1098.

Perreira, E. A., Green, A., Nandi, D., & Aziz, T. Z. (2007). Deep brain stimulation: Indications and evidence. *Expert Review Medical Devices 4*, 591-603.

Pett, M., Lackey, N., & Sullivan, J. (2003). *Making Sense of Factor Analysis.* Thousands Oaks: Sage Publications.

Phelps, R. (2008). *The role and importance of standardized testing in the world of teaching and training.* Marrakesh, Morocco: Cadi Ayyad University.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon, 9 (5)*, 1-6.

Prensky, M. (2003). *Digital game based learning.* New York: Games2Train.

Price, C. (2009). Why don't my students think I'm groovy? The new R's for engaging millenial learners. *The Teaching Professor, 23*, 1-6.

Public School Review. (2018, Febuary 15). *Public School Review*. Retrieved from Public School Review: https://www.publicschoolreview.com/blog/public-school-boards-demystified-how-parents-can-influence-the-boards-decisions

Queens University of Charlotte. (2019). *Benefits of diversity in schools.* Charlotte, NC: Queens University of Charlotte.

Rae, G., Cork, J., Karpinski, A., Farris, H., & Swartz, W. (2015). Using the brain we have: Dissection of the human brain assists the medical student in learning and retaining neuroanatomy. *The FASEB Journal 29 (1)*.

Rand Corporation. (2012). *Tests and the teachers: What student achievement tests do and don't tell us about teacher effectiveness.* Santa Monica, CA: Rand Coporation.

Rastogi, R., & Silver, E. (2014). *Association of Music with Stress, Test Anxiety, and Test Grades among High School Students.* Scarsdale.

Rasul, S., & Bukhsh, Q. (2011). A study of factors accecting students performance in examination at university level. *Procedia- Social and Behavioral Sciences 15*, 2042-2047.

Ray, B. (2018, April 20). *National Home Education Research Institute*. Retrieved from NHERI: https://www.nheri.org/homeschool-population-size-growing/

Response Media. (2017). *Generation Z the newest generation* . Atlanta, GA: Response Media.

Reynolds, B. (2018, January). *Getting Smart.* Retrieved from Word Press: https://www-gettingsmart-com.cdn.ampproject.org/v/www.gettingsmart.com/2018/01/how-does-project-based-learning-prepare-generation-z-for-the-future

Richmond, V., & McCroskey, J. (1995). *Nonverbal behaviors in interpersonal relations (3rd Ed.).* Boston, Ms: Allyn & Bacon.

Rietveld, T., & Hout, R. (1993). *Statistical techniques for the study of language and language behaviour.* Nijmegen, Netherlands: Radboud University.

Ritz, P., & Berrut, G. (2005). The importance of good hydration for day to day health. *Nutrition Reviews*, 6-13.

Rogers, K. (1991). *Realtionship of grouping practices to the education of the gifted and talented learner.* St. Paul, MN: The National Researcher Center of the Gifted and Talented.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., . . . Tenopil, C. (2008). The Google generation: The inoformation behavior of the researcher of the future. *Aslis Proceedings 60(4)*, 2090-310.

Royston, P., Altman, D., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat Med 15 (25)*, 127-141.

Rubin, B. (2008). Detracking in context: How local constructions of ability complicate equity geared reform. *Teachers College Record 110 (3)*, 646-696.

Rubin, H., & Rubin, I. (2012). *Qualitative interviewing. The art of hearing data (3rd).* Thousand Oaks, CA: Sage.

Rumberger, R., & Palardy, G. (2005). Scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Education Research Journal 42(1)*, 3-42.

Russell, G., McCown, E., & Broyles, J. (1985). Effects of noise on sensitivity of introverts and extraverts to signals in a vigilance task. *Personality and Individual Differences Vol. 6 (2)*, 237-241.

Rusticus, S., & Lovato, C. (2014). Impact of sample size and variability on the power and type I errors of equivalence tests. *Practical Assessment Research and Evaluation 19 (11)*, 1-10.

Salend, S. (2012). Teaching students not to sweat the test. *Phi Delta Kapan 93 (6)*, 20-25.

Sambell, K., McDowell, L., & Brown, S. (1997). But is it fair?: An exploratory study of student perceptions of consequential validity of assessment. *Studies in Educational Evaluation 23(4)*, 349-371.

Santamaria, L. (2009). *Culturally responsive differentiated instruction: Narrowing gaps between best pedagogical practices benefitting all learners.* Teachers College Record.

Sawyer, R. (2010). *The usefulness of high school average and ACT scores in making college admission decisions.* Iowa City, IA: ACT Incorporated.

Schanzenbach, D. (2012). *Limitations of experiments in educational research.* Evanston, IL: Institute for Policy Research at Northwestern University.

Schawbe, L., & Wolf, O. (2009). The context counts: Convergent learning and test environments prevents memory retrival impairment following stress. *Cognitive, Affective, and Behavioral Neuroscience 9(3)*, 229-236.

Schechtman, E., & Yitzhaki, S. (2010). Identifying monotonic and non-monotonic relationships. *Economic Letters 116(1)*, 23-25.

Scheetz, G. (2000). Peoria. In E. Callary, *Place names in the Midwestern United States.* Lewistown, NY: Mellen Press.

Schoenherr, N. (2006). *Discovering why study groups are more effective.* St. Louis, MO: Washington University.

Schram, T. (2006). *Conceptualizing and Proposing Qualitative Research.* Columbus, OH: Pearson.

Schumaker, R., & Lomax, R. (2016). *A beginner's guide to structural equation modeling.* New York City: Taylor & Francis.

Secher, M., & P, R. (2012). Hydration and cognitive performance. *The Journal of Nutrition, Health, and Aging*, 325-329.

Segers, M., Dochy, F., & Cascallar, E. (2003). *The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment.* Dordecht: Kluwer Academic Publishers.

Seigel, D. (2016, February 16). *ACT, ACT admission requirements: The importance of the ACT*. Retrieved from Prep Scholar Web Site: http://blog.prepscholar.com/act-admission-requirements-importance-of-the-act

Sevier, F. (1957). Testing the assumptions underlying multiple regression. *The Journal of Experimental Education, 25 (4)*, 323-330.

Sheer, S. (2011). *The Effect of Background Music, Speech, and Silence on Office Workers Selective Attention.* Tallahassee: Florida State University.

Siegel-Hawley, G. (2012). *How non-minority students also benefit from racially diverse schools.* Washington, DC: National Coalition on School Diversity.

Small, G., Moody, T., Siddarth, P., & Brookheimer, S. (2009). Your brain on Google:Patterns of cerebral activation during internet searching. *The American Journal of Geriatic Psychiarty*, 116-126.

Smith, S. (1985). Background music and context dependent memory. *American Journal of Psychology 98*, 591-603.

Smith, S., & Vela, E. (2001). Environmental context dependent memory: A review and meta analysis. *Pyschonomic Bulletin and Review 8*, 2003-220.

Smyth, E. (2010, June). Single-Sex education: What does research tell us? *Revue Francaise de Pedagogie 171*, pp. 47-55.

Sparks, S. (2015, January 28). Differentiated instruction: A primer. *Education Week*.

Sparrow, B., Liu, J., & Wegner, D. (2011). Google effects on memory:Cognitive consequences of having inforamtion at our fingertips. *Science 333*, 776-778.

Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology 3*, 271-295.

Spielberger, C., & Vagg, R. (1995). *Test Anxiety.* Bristol: Taylor and Francis.

Srivastava, P., & Hopwood, N. (2009). *A practical iterative framework for qualitative data analysis.* Alberta: Sage Publications.

Stack, L. (2018, March 1). Are you 21-37? You might be a millenial. *New York Times*.

Standing, L., Boisvert, K., Dayholos, K., & Gagnon, A. (2008). People, clothing, music, and arousal as contextual retrieval cues in verbal memory. *Perceptual and Motor Skills (107)*, 523-534.

Steiger, J. (1990). Structural model evaluation and modification:An interval estimation approach. *Multivariate Behavioral Research 25*, 173-180.

Sternberg, R. (2006). *Cognitive Pyschology (4th).* Belmont, CA: Thomson Wadsowrth.

Stevens, J. (2009). *Applied multivariate statistics for the Social Sciences 5th.* New York, NY: Taylor & Francis.

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan 72 (7)*, 534-539.

Struyven, K., Dochy, F., & Janssesn, S. (2005). Students' perceptions about evaluation and assessment: A review. *Assessment and Evaluation in Higher Education 30 (4)*, 325-341.

Studenmund, A., & Cassify, H. (1987). *Using econometrics: A practical guide.* Boston: Pearson.

Subban, P. (2006). Differentiated instruction: A research basis. *Interantional Education Journal 7(7)*, 935-947.

Subedi, D. (2016). Explanatory sequential mixed method design as the third research community of knowledge claim. *American Journal of Educational Research, 4 (7)*, 570-577.

Sweeney, R. (2006). *Millennial behaviors and demographics.* Newark, NJ: New Jersey Institute of Technology.

Szinnai, G., Arnaud, M., Linder, L., & Keller, U. (2004). Effect of water deprivation on cognitive-motor performance in healthy men and women. *American Journal of Physiology-Regulatory Intergrative and Comparative Physiology*, 325-329.

Tabachnick, B. G., & Fiddell, L. S. (2001). *Using multivariate statistics.* Boston: Allyn & Bacon.

Tanzer, U., von Fintel, A., & Eikerman, T. (2009). Chewing gum and concetration performance. *Psychological Reports 105*, 372-374.

Tapscott, D. (1998). *Growing up digital: The rise of the net generation.* New York: McGraw-Hill.

Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world.* New York: McGraw- Hill.

Tari, A. (2010). *Az Y és mögötte a Z generáció* . Budapest: Budapest University. Retrieved from Budapest University.

Taylor, J. (1999). *An introduction to error analysis: The study of uncertainties in physical measurements.* South Orange, NJ: University Science Books.

Teeter, P., & Sandberg, J. (2016). Constraining or enabling green capibility development? How policy uncertainty affects organizational responses to flexible environmental regulations. *British Journal of Management*.

The Center for Generational Kinetics. (2017). *The state of Gen Z 2017: Meet the throwback generation.* Austin, TX: The Center for Generational Kinetics.

The College Board. (2016). *The College Board*. Retrieved June 13, 2016, from The College Board Website: https://collegereadiness.collegeboard.org/psat-nmsqt-psat-10

The College Board. (2017). *AP bulletin for students and parents.* New York, NY: The College Board.

The College Board/National Merit Scholarship Corportation. (2016). *Supervisor's Manual PSAT/NMSQT.* Princeton, NJ: The College Board.

The Kapersky Lab. (2015). *The rise and impact of digital amnesia.* Moscow: Kapersky Lab.

Thompson, B. (2010). *Exploratory and confirmatory factor analysis.* Washington DC: 2010.

Thorne, S., Reimer Kirkham, S., & MacDonald-Emes, J. (1997). Focus on qualitative methods: Interprative Description. *Research in Nursing and Health 20*, 169-177.

Tindell, D., & Bohlander, R. (2012). The use and abuse of cell phones and test messaging in the classroom: A survey of college students. *College Teaching 60 (1)*, 1-9.

Togut, T. (2018, September 6). *EDEX 790*. Retrieved from Archive: https://web.archive.org/web/20090111075836/http://www.ed.sc.edu/caw/definition.html

Tomlinson, C. (1999). Mapping a route toward a differentiated instruction. *Educational Leadership 57 (1)*, 12.

Tomlinson, C. (2001). *How to differentiate instruction in mixed-ability classrooms (2 ed.).* Alexandria, VA: Classrooms.

Tomlinson, C. (2015, January 27). Differation does, in fact work. *Education Week*.

Tomprowski, P., Davis, C., Miller, P., & Naglieri, J. (2008). Exercise and children's intelligence, cognition, and academic achievement. *Educational Phsychology Review 20(2)*, 111-131.

Tulving, E., & Thompson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review 80*, 352-373.

Turner, A. (2015). Generation Z: Technology and social interest. *Journal of Individual Psychology 71(2)*, 103-113.

Twenge, J. (2017, August 15). How the smartphone affected an entire generation of kids. *The Conversation*. Retrieved from The Conversation.

Vaez, M., & Laflamme, L. (2008). Experienced stress, psychological symptom, self-rated health, and academic achievement: A longitudinal study of Swedish university students. *Social Behavior and Personality*, 183-196.

Valero-Cuevas, F., Sklaar, S., & Peters, M. (2019, January 15). Standardized tests unfairly favor wealthier students in college admissions. *Sacramento Bee*.

Van de Watering, G., Gijbels, D., Dochy, F., & Van de Rijt, J. (2008, December). Students' assessment preferences, perceptions of assessment and their relationship to study results. *Higher Education, 56*, 645-658.

Van Merrienboer, J., & Kirschner, P. (2013). *Ten steps to complex learning (2nd Rev.ed).* New York, NY: Taylor and Francis.

Vanhout-Wolters, B., Simons, P., & Volet, S. (2000). Active learning;Self-diected learning and independent work. *New Learning*, 21-36.

Veen, W., & Vrakking, B. (2006). *Homo Zappiens:Growing up in the digital age.* London: Network Continuum.

Veuglers, W. (2004). Between control and autonomy.Restructuring secondary education in the Netherlands. *Journal of Educational Change 5*, 141-160.

Vogel, I., Verschure, H., van der Ploeg, C., Burg, J., & Raat, H. (2009). Adolescents and MP3 Players: Too Many Risks, Too Few Precautions. *Pediatrics*, 953-958.

Wahlstrom, K. (2002). Changing times: Findings from the first longitudinal study of later high school start times. *NASSP Bulletin*.

Walkers, G. (2004, December 3). *Northland Sensory Insights*. Retrieved from Matrix Sciences: http://www.northlandsensoryinsights.com/measuring-responses-for-acceptance-and-preference-testing/

Wang, X. (2004). *Chinese EFL students perceptions of classroom assessment environment and their goal orientations in the college English course.*

Wechsler, H., Devereaux, R., Davis, M., & Collins, J. (2000). Using school environment to promote physical activity and healthy eating. *Preventative Medicine 31*, 121-137.

Weinstein, C. (1979). The physical environment of school: A review of the research. *Review of Educational Research 49 (4)*, 577-610.

Wiersma, W., & Jurs, S. (2009). *Research Methods in Education.* Boston: Pearson.

Wignall, A. (2017, January 27). *College Raptor*. Retrieved from College Raptor: https://www.collegeraptor.com/getting-in/articles/questions-answers/what-is-an-act-sat-superscore/

Williams, A. (2015, September 18). Move over Millennials, here comes Generation Z. *New York Times*.

Wolf, M. (2008). *Proust and the squid:The story and science of the reading brain.* New York, NY: Harper Collins.

Wolters, C. (2003). Regulation of motivation:Evaluating an underemphasized aspect of self-regulated learning. *Educational Pyschologist 38(4)*, 189-205.

Xie, Y. (2012). Values and limitations of statistical models. *Res Soc Stratif Mobil 29 (3)*, 343-349.

Zhang, H., Miller, K., Cleveland, R., & Cortina, K. (2018). How listening to music affects reading: Evidence from eye tracking. *Journal of Experimental Pyschology: Learning, memory, and cognition 44 (11)*, 1778-1791.

Zhu, J., & Han, L. (2011). Main factors effecting reliability of test papers. *Journal of Languge Teaching and research 2 (1)*, 236-238.

Zimmerman, B. (2002). Achieving academic excellence:A self-regulatory perspective. In M. (. Ferrari, *The pursuit of excellence through education* (pp. 85-110). Mahaw, NJ: Erlbaum.