QUANTIFYING PERFORMANCE:

The Creation and Examination of an Index of High School Grade Point Average and SAT Score to Predict College Performance

by

Todd Roecker Wadsworth

Liberty University

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

Liberty University

November, 2016

QUANTIFYING BOYS' UNDERPERFORMANCE:

INDEXING HIGH SCHOOL GRADE POINT AVERAGE AND SAT SCORE

by Todd Roecker Wadsworth

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

Liberty University, Lynchburg, VA

November, 2016

APPROVED BY:

Amanda J. Rockinson-Szapkiw, Ed.D., Committee Chair

John C. Bartlett, Ed.D., Committee Member

Richard J. Cleary, Ph.D., Committee Member

ABSTRACT

A variable derived from commonly available performance metrics was created. The derived variable was an index created from the quotient of each student's high school cumulative grade point average and SAT composite test score. Its predictive validity for college performance of both first-semester males and females was examined. The data used in the study was archival and obtained from a college freshmen cohort of 544 students. The analysis was carried out by conducting three separate bivariate correlation analyses. Descriptive statistics of the index were also reported, both as a whole and disaggregated by sex. The descriptive statistics included means, variances, and graphical views. The importance of this study was to provide a different quantitative perspective on performance and to examine whether and how much that different measure predicts performance.

Keywords: sex, males, females, performance, underperformance, cognitive, noncognitive, SAT, GPA, standardized tests

Acknowledgements

I need to acknowledge my deep gratitude for all the support I have received in completing this project. First, thanks to my committee of Dr. Rockinson-Szapkiw, Dr. Cleary, and Dr. Bartlett. Without their guidance, patience, and meticulous feedback I could never have finished. I also thank Dr. Keith who got me off to a good start with her positive attitude on seemingly everything. I thank all my students and players from whom I have learned much more than I have taught. I acknowledge and thank my many teachers, coaches, and mentors for their examples and lessons that provided me with the courage and confidence to pursue something like this. I thank my mother and father, Barbara and John, who gave of themselves without limit and provided the best childhood and trainings any boy could ever have. I thank my many family members, past and present, who have contributed so much to my life and my capabilities. I thank my son, Benjamin, with whom I have had numerous spirited discussions about education. He has been a source of inspiration and love for me for 24 years. Last, but surely not least, I thank my wife, Beverly, who has only been with me a couple years, but who has been a heroic example of so many things and who has provided unflinching understanding, support, and comfort. God put all these amazing people along my path and I am humbled with gratitude.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACT composite (ACTC)

Armed Services Vocational Aptitude Battery (ASVAB)

Educational Testing Service (ETS)

First-year college grade point average (FYGPA)

First-semester college grade point average (FSGPA)

Grade point average (GPA)

High school grade point average, cumulative (HSGPA)

Individual Performance Index (IPI)

Institutional Review Board (IRB)

National Association for College Admission Counseling (NACAC)

National Collegiate Athletic Association (NCAA)

National Council for Education Statistics (NCES)

Programme for International Student Assessment (PISA)

SAT composite (SATC)

Socio-economic status (SES)

Third International Mathematics and Science Study (TIMSS)

Note: In 1996, ACT was an acronym for American College Testing. The company dropped the acronym and has since been known as ACT. In 2004, the College Board changed the name of the Scholastic Aptitude Test to the SAT.

CHAPTER ONE: INTRODUCTION

Introduction

College and university personnel advise and admit students using information that predicts college success. Predicting college success is vital for individuals, for academic institutions, and for the country. For students, successfully obtaining a degree "correlates strongly with most important social and economic outcomes such as economic success, health, family stability, and social connections" (Hout, 2012, p. 379). Conversely, for academic institutions, attrition is costly. Failure to admit students who will successfully obtain degrees can result in loss of accreditation and finances for the institution and also the state if institutions are public. The percentages of non-graduates of public institutions range from 26.2% in Iowa to 68.5% in Alaska with annual costs to the states ranging from tens of millions of dollars to over $1 billion in California and Texas (State Higher Education Executive Officers, 2012). With attrition rates ranging from 30-70% and the high cost of attrition, it is critical that colleges and universities have a metric that accurately predicts student success and can be used to advise students in their academic careers.

Standardized test scores and high school grades are frequently used as predictors of college success, grades, and retention (Bai, Chi, & Qian, 2014; Bettinger, Evans, & Pope, 2013). But, these measures account for approximately 25% of the variation in college grades (Sparkman et al., 2012). Scott-Clayton (2012), who examined the validity issues of standardized tests, demonstrated that high error rates exist. In order to make better admittance and advising decisions, higher education personnel need more accurate measures such as interpreting the tests differently, using supplemental tests, or improved metrics. Examining the predictive ability of different measures and disaggregating by selected demographics could reveal important insights and improve predictions. Disaggregation of data based on sex is sometimes neglected even in well-known compilations, although performing this disaggregation may prove beneficial. The Condition of College and Career Readiness report by ACT (2015), in a detailed 20 pages of statistics that disaggregates by ethic group and numerous other ways, does not disaggregate by sex, even though the same report lists "interpersonal, self-regulatory, and task-related behaviors" (p. 3) as important for educational success and other reports indicate that males and females as groups differ in these characteristics (Dash, 2011; Casillas et al., 2012; Ingalhalikara et al., 2014).

Background

Research and the media from Western nations have frequently reported that males are underperforming with respect to their female peers (Driessen & van Langen, 2013; Jonsson, 2014; Klevan, Weinberg, & Middleton, 2016; Kutner et al., 2007; Mathews, 2009). Boys "are struggling in school, with lower grades, more discipline problems, more learning disabilities, and more behavioral disorders than girls" (King & Gurian, 2006, p. 56). Van Houtte (2004) writes that "in recent years, in many countries, increasing attention has been paid to the underachievement of boys in comparison with girls" and that it has been "demonstrated repeatedly that in general girls outperform boys" (p. 159). Vincent-Lancrin (2008) reports that from 1974 to 2003, "the gap in academic levels between boys and girls aged 16 widened in favour of girls, at the aggregate level … in mathematics (with the girls catching up) and in English (where the gap widened)" (p. 284).

Males are less engaged in school and experience higher frequency of behavioral issues including truancy, disruptions in school, lack of effort on academic assignments, and criminal activities (Gurian & Stevens, 2005; Juelskjær, 2008). These authors point out behavioral issues and lack of engagement as important factors in males' underperforming. Grades, a common measure of academic performance, are influenced by ability, but also by numerous other factors such as behavior, motivation, goal-setting, grit, and self-control (Farrington et al., 2012). Boys tend to have lower grades than girls.

However, direct comparison of ACT and SAT scores, often considered indicators of academic ability rather than performance (Coyle & Pillow, 2008; Koenig, Frye, & Detterman, 2008), indicate that males score moderately higher than females toward the end of high school (ACT, 2013c; SAT, 2013b). Standardized tests scores are highly correlated with IQ and grades are highly influenced by noncognitive factors.

Both standardized test scores and high school GPA (HSGPA) are often used as predictors of college success. Due to its relationship to cognitive competence, high school GPA can be a useful measure for college success (Curie et al., 2012; deAngelis, 2003; Gaertner & McClarty, 2015; Meriac, 2012). The predictive validity of ACTC and SATC scores are also studied extensively (Gaertner & McClarty, 2015). Unfortunately, examining standardized tests or high school grades independently or in linear combinations does not take into account a student's performance relative to a student's ability. If a student is in the 70^{th} percentile in ability but in the 30^{th} percentile in performance, or vice versa, this apparent discrepancy is not measured or accounted for by either grades or test scores alone. This study creates an index that calculates ability relative to performance and uses the index to predict college performance for males and females.

Problem Statement

"SAT, ACT, and HSGPA are the most heavily researched and relied upon college-readiness indicators in the United States" (Gaertner & McClarty, 2015, p. 22). Standardized test scores are a common measure of ability more than performance (Coyle & Pillow, 2008; Koenig, Frye, & Detterman, 2008). ACTC and SATC scores show males and females with statistically similar standardized test scores (ACT, 2013c; SAT, 2013b). HSGPA is common measure of performance. Girls earn higher grades than boys in all major subjects, including science and math, throughout elementary, middle, and high school (Cornwell, Mustard, & Van Parys, 2013). By one measure, boys have at least equal potential, but by another measure (grades) they are not performing as well as girls. What is not addressed by either measure alone is performance relative to ability. That is, a HSGPA of 3.0 may seem either excellent or merely average, depending on the particular student's ability.

Purpose Statement

The purpose of this study was to create an index that considers students' academic performance relative to their ability. The index was created from the quotient of a student's cumulative high school grade point average (HSGPA) and the student's composite SAT score (SATC). This predictor variable was referred to as the Individual Performance Index (IPI) and was used to predict college performance as measured by first-semester college grade point average (FSGPA). Since the predictor and criterion variables are both continuous, and the expected relationship is one predictor and one criterion, a simple bivariate correlation model is used. The study also examined how well the IPI predicts the criterion variable separately for males and females.

Significance of the Study

While ACT scores, SAT scores, and HSGPA have been studied extensively as measures of high school performance and predictors of college success (Bai, Chi, & Qian, 2014; Bettinger, Evans, & Pope, 2013; Crede et al., 2010; Ledsema & Obukova, 2015), the Individual

Performance Index (IPI) provides a measure that considers grades relative to ability, rather than just looking at grades alone. If the IPI is found to be a useful tool for predicting college success, it can be used in a variety of ways to inform college admittance and advising procedures. Students with grades lower than expected, for example, moderate GPA but high SATC scores, might be counseled that even moderate success in college may require more effort because the curriculum will be more challenging than it was in high school. The primary intended audience for this study is admissions and student affairs personal.

Research Questions

The research questions guiding this study examine the predictive validity of the IPI.

RQ$_{1:}$ Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average?

RQ$_{2:}$ Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males?

RQ$_{3:}$ Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females?

Null Hypotheses

The null hypothesis for this study are:

H$_{01}$**:** There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average.

H$_{02:}$ There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males.

H_{03}: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females.

Identification of Variables

Predictor Variable

The predictor variable was a derived variable created from two measures commonly used in the college admissions process (Ledsema & Obukova, 2015). The first measure was cumulative High School GPA (HSGPA) as reported to the college on the application form. The second was the SAT composite (SATC) test score. HSGPA and SATC are among the most widely used metrics in college admissions to predict college success (Balsa, Guiliano, & French, 2011; Schmitt et al., 2009). The SATC test is primarily a measure of ability (Brown et al., 2008; Conard, 2006; NCES, 2011; Toomela, 2008). Grades are influenced by ability, but are also influenced by myriad noncognitive factors and "measure a student's ability to "get it done" in a more powerful way than do SAT scores" (Bowen et al., 2009, pp.123-124). The derived variable was the quotient of the percentile rank of the student's HSGPA over the percentile rank of their SATC score. This created an index that indicated whether a student's grade performance was at the same percentile corresponding with that student's ability. The index was referred to as the Individual Performance Index (IPI).

Criterion Variable

The criterion variable was the first-semester college GPA (FSGPA). This was measured on a 4.0–5.0 scale by the college in the study. Grades that were reported on a 100-point scale were converted to the 4.0–5.0 scale.

Additional Variable

Sex also served as a variable. The primary purpose of this study was to examine the predictive value of the IPI for college grades. It was also to examine whether the IPI's predictive value varied when examined separately for each sex.

Definitions

Some of the following terms are imprecise because they refer to phenomena that are inherently ineffable. Nonetheless, these terms and the phenomena they represent are essential to education and their somewhat nebulous character is grounds for explication or acknowledgement, not avoidance. It is important to define them as they as central to this study.

Ability or aptitude - A measure or quantification of individual mental capacities directly involved in educational endeavors and which impact success (Harackiewicz, Barron, Tauer, & Elliot, 2002; Komarraju, Ramsey, & Rinella, 2013).

Cognitive - This refers to intellectual processes such as mathematical reasoning, verbal skills, vocabulary, and working memory (Willingham, 2013).

Noncognitive - This refers to skills or factors like perseverance, attitude, grit, self-regulation, self-efficacy, and curiosity (Farrington et al., 2012; Tough, 2012).

Quasi-Cognitive - Factors which lie between noncognitive and cognitive factors; that is, they include elements of both. For example, creativity, emotional intelligence, and confidence (Kyllonen, Walters, & Kaufman, 2011).

Gender vs *Sex* vs *Gender-Identity*. It is outside the scope of this study to treat with how the terms gender, sex, and gender-identity are, or should be, used. Usage is not consistent and the binary regimes and heteronormativity are challenged in any case (Brubaker, 2016). For the purposes of this study, the terms sex and gender will be used to signify the sex/gender that an individual student reports on the college application. On the Common Application website, www.commonapp.org, which is used by nearly 700 colleges including the study college here, the required question is phrased as "sex assigned at birth" although there is also a text field for additional student input on "gender identity."

CHAPTER TWO: LITERATURE REVIEW

Introduction

This chapter reviews the research that grounds this study in existing conceptual frameworks. The first portion of the chapter includes general theoretical frameworks that together form the support and context in which the study takes place. These frameworks inform, guide, and provide background for the research and topics directly related to this study.

The second part of this chapter provides a review of the topics and literature relevant to this particular study. The literature review will examine cognitive and noncognitive frameworks as they relate to measures of academic performance, sex differences against this backdrop, male underperformance in academics, and the role of indexing in developing meaningful quantitative conclusions. This research guides the formation of the hypotheses and the conduct of the study. Finally, the chapter summarizes this background information, and provides the basis for the conduct of this research.

Theoretical Frameworks

The primary framework in this study is that of indexing. Indexing theory provides no single major theorists, aside perhaps from philosophers of science who consider what numbers represent or the concept of quantity. However, the framework of indexing has a long, distinguished, and ubiquitous history (Fischer, 1923). Contributions to indexing theory come from myriad perspectives and are surprisingly multi-faceted and robust. A few examples are Cohen's *d*, *z*-scores, any percentage, business ratios, and IQ. These are all indices. An indexing framework is used here to improve the practical interpretation of scores and to provide a quantitative view of the qualitative variables in the literature. The literature on indexing will support the methods used here to improve the interpretation of traditional measurements of performance and ability, or cognitive and noncognitive constructs. Indexing is used to make raw numbers more meaningful by providing context. For example, the number five by itself is neither big nor small. When compared to 20,000 it is relatively small; when compared to 0.01 it seems big. An index can be used to give context (Cohen's *d*), make more informative comparisons (return on assets), remove artifacts (inflation), track national progress and changes (Human Development Index), or quantify phenomena that resist numerical analysis (Work Problems Index, Marital Conflict Index, Depression Index).

This study applies the framework of indexing to high school grades in order to gain information about academic performance. It is not immediately clear whether a 3.0 GPA should be viewed as performing as expected, underperforming, or overperforming. This will depend somewhat on the student's ability. If a student is at the 95th percentile in aptitude, it could be expected the student's grades should be at approximately the 95th percentile. If a student is 50th percentile aptitude, a grade point average in the 90th percentile implies that the student is performing well above what might be expected. While there is disagreement about what grades measure, even what they should measure (Allen, 2005), there is little disagreement they are the primary measure of performance in high school and college. "Despite problems with grading reliability and disciplinary and institutional grading differences, [GPA] is still the most widespread performance measure" (Robbins et al., 2004, p. 262). This index provides a measure of performance relative to a student's ability, not simply performance relative to other students.

The index in this study is constructed by taking the percentile of the HSGPA and dividing by the percentile of the SATC scores. Standardized test scores, including the SAT, are frequently referred to as measures of ability (Brown et al., 2008; Conard, 2006; NCES, 2011; Toomela, 2008). They are also referred to as cognitive or aptitude tests (Komarraju et al., 2013).

The reliability and validity of the SATC is well-documented (College Board, 2010). The validity of grades is demonstrated by wide-spread reliance on it, skeptics not withstanding (Balsa, Guiliano, & French, 2011). The index, by its construction, inherits its validity as a measure of performance relative to ability.

The second framework is the emerging research focusing on the importance of noncognitive factors in educational success (Gaertner & McClarty, 2015; Lipnevich & Roberts, 2012). This is a broad framework because noncognitive factors cover a range of subtopics, but these noncognitive factors have been shown to have substantial correlations with and influence on grades. Success in college can be attributed to both cognitive and noncognitive factors, and noncognitive factors influence grades more than tests of ability (Duckworth, Quinn, & Tsukayama, 2012). The study of noncognitive factors is growing (Sohn, 2010). Economists and psychologists are starting to realize that noncognitive factors highly influence success, even more than IQ (Tough, 2012). Noncognitive factors are more under the control of students, as illustrated by the common admonition to students to get their grades up, while not being told to get their IQs up. The immutability of cognitive test scores is supported when efforts to improve noncognitive behaviors do not always improve cognitive results (Holmlund & Silva, 2014).

Ancillary to this and the reason for examining males and females separately in this study are the documented discrepancies between the sexes in noncognitive characteristics and behaviors (Koul, Roy, & Lerdpornkulrat, 2012; Weis, Heikamp, & Trommsdorff, 2013). Girls' math self-efficacy is not as high as boys (Hyde, 2014). Female students are more likely than males to attribute low marks to lack of ability or task difficulty (McClure et al., 2011). Females are dominating in college enrollments at nearly every level (Klevan, Weinberg, & Middleton, 2016). Because of sex differences in noncognitive factors, any analysis which brings noncognitive factors into play should disaggregate to reveal informative differences between sexes. A lesson that can be derived from Gilligan's (1982) criticism of Kohlberg's (1976) framework for ethical development is that omission of sex considerations is sometimes an Achilles heel in the framework (Bussey & Bandura, 1999). "Boys are biologically, developmentally, and psychologically different from girls--and teachers need to learn how to bring out the best in every one" (Tyre, 2006, p. 1). Frameworks that support or account for these differences are helpful.

Literature

> 'Non-cognitive factors' is a misleading but entrenched catch-all term for factors such as motivation, grit, self-regulation, social skills. . . in short, mental constructs that we think contribute to student success, but that don't contribute directly to the sorts of academic outcomes we measure, in the way that, say, vocabulary or working memory do (Willingham, 2013, p. 1)

The research distinguishes between cognitive and noncognitive factors and how they impact student performance. Cognitive and noncognitive are analytically exhaustive terms, but are not independent and seem to evade precise definition. Even in Willingham's statement, above, "non-cognitive factors" are referred to as "mental constructs" – an apparent contradiction. These terms are of central importance to this study. A precise delineation is not found in the literature, and will not appear here, however some clarification about usage is warranted.

Cognitive and Noncognitive Factors

Identifying and measuring noncognitive skills is an emerging and important topic. Noncognitive skills have been identified as predictors of academic success in the past (McDaniel, Halter, & Hartford, 1961; Kipnis, 1962; Okun 1980; McGanney & Ganoo, 1995;

Messick, 1979; Tracey & Sedlacek, 1984). However, there is not a lot of research directed toward these as predictors (Thomas, Kuncel, & Credé, 2007). Noncognitive factors have been included in research, but are now being examined more closely in education research (Lipnevich & Roberts, 2012). Sohn (2010) writes that an "emerging body of literature persuasively argues that noncognitive skills are as important as or even more important than cognitive skills" (p. 125). Economists and psychologists used to believe the critical factor in a child's success was IQ, but they are coming to realize that success is heavily impacted by noncognitive factors (Tough, 2012).

Noncognitive factors are becoming the basis for changing educational policy. Lipnevich and Roberts (2012) note,

> the growing role of noncognitive factors in large scale international assessments with an attendant impact on education and economic policy (e.g., PISA, Naemi et al., in press) and even legislation…. In fact, in countries as diverse as the United States, United Kingdom, Finland, Korea, Israel, and Singapore, noncognitive skills have been elevated to playing a central role in national curricula. (p. 173)

While the importance of noncognitive factors is gaining recognition and the division of factors into cognitive and noncognitive seems straightforward, the distinction is nonetheless problematic (Messick, 1979). Farrington et al. (2012) write that "“noncognitive” [is] an unfortunate word. It reinforces a false dichotomy between what comes to be perceived as weightier, more academic “cognitive” factors and what by comparison becomes perceived as a separate category of fluffier “noncognitive” or “soft” skills" (p. 2).

In "Non-cognitive Skills: Bad Name, Really Important," Van Ark (2012) acknowledges the importance of noncognitive factors in success in college and work, but refers to them as "so called" noncognitive skills. Categorizing precisely what constitutes noncognitive versus cognitive skills proves difficult for the specific reason that virtually all aspects of human behavior are predicated on cognition to some degree (Borghans, Duckworth, Heckman, & Weel, 2008).

Farrington et al., (2012) list five general categories of noncognitive skills including academic behaviors, academic perseverance, academic mindsets, learning strategies, and social skills. Tough (2012) writes about a set of strengths that predict "life satisfaction and high achievement" (p. 76). These include grit, self-control, zest, social intelligence, gratitude, optimism, and curiosity. Mayer, Roberts, and Barsade (2008) write that emotional intelligence is a noncognitive skill (this reinforces the problems inherent in separating noncognitive from cognitive.). Tracey and Sedlacek's (1984) often-cited Non-Cognitive Questionnaire (measures positive self-concept, realistic self-appraisal, dealing with racism, preference for long-range goals, availability of a strong support person, successful leadership experience, demonstrated community involvement, and knowledge acquired in a field.

Another term encountered in this emerging discussion is metacognitive. It is not clear a useful or accurate distinction is being made here, but the skills referenced in the metacognitive discussions are relevant to grades and standardized test scores. Noncognitive skills are sometimes referred to as metacognitive. The Washington College Access Network (n.d.) refers to "noncognitive variables (also called “meta cognitive learning skills”)" in its training. Rahman and Masrur (2011) view metacognition as “thinking about thinking” and as a "cognitive skill which involves…the monitoring of comprehension, problem solving and other cognitive skills" (p. 135). Shimamura (2000) refers to metacognition as the "evaluation and control of one’s cognitive processes" (p. 313).

Conley (2013) proposes that "metacognitive" is a better term than noncognitive. He describes "what has previously fallen under the label of noncognitive factors as 'metacognitive learning skills'.…metacognitive in this context includes all learning processes and behaviors involving any degree of reflection, learning-strategy selection, and intentional mental processing" (p. 20). The gap Conley leaves open are student actions that do not require self–reflection, selection, or processing. For examples, grit, attendance, and confidence. Metacognition, as a conscious "practice," is classified here as a noncognitive skill (that involves cognition), somewhat along the lines of a specific learning strategy.

> It is clear that simple contrasts such as cognitive versus noncognitive are popularly embraced in spite of the dangers of stereotyping, probably because they highlight major distinctions worth noting. It is in this spirit, then, that some major features of cognitive and noncognitive assessment are addressed—with an insistence that cognitive does not imply only cognitive and that noncognitive does not imply the absence of cognition. (Messick, 1979, p. 282)

"The importance of metacognition in the process of learning is an old idea that can be traced from Socrates' questioning methods to Dewey's twentieth-century stance that we learn more from reflecting on our experiences than from the actual experiences themselves" (Tanner, 2012, p. 113). The distinction is whether a thinker is capable of standing sufficiently "outside" their own thinking to consider their thinking objectively, or whether metacognition is used simply to imply we should think about how we go about learning. This second sense is how it is used in the educational literature and how it will be used here. Perhaps "metalearning" is a better term, but metacognition is the term that will be used here.

The Educational Testing Service (ETS) uses the term Quasi-Cognitive Factors, which are often "considered somewhere in between cognitive and noncognitive factors. They may be measured with performance tests, or ability tests, but they also reflect affective qualities" (Kyllonen, Walters, & Kaufman, 2011, p. iii). Kyllonen et al. list creativity, emotional intelligence, metacognition and confidence, and cognitive style as quasi-cognitive factors. Their report uses an example of metacognition as whether students can predict if they will know how to answer an exam question.

Measuring and predicting performance. Some noncognitive factors are easy to isolate and measure, for example attendance, hours spent on work, and persistence, as measured by retention. Inroads are being made in measuring less precise noncognitive qualities, including things like "personality constructs" (Lounsbury et al., 2003, p. 1231). It had been a safer domain to restrict measurement to objectively observable performance, for example, whether a student can factor a particular quadratic equation or read a passage and answer questions about it. This safe harbor is giving way to examination of less objective and less clearly defined phenomena. Conley (2013) writes that "perhaps it's time to move beyond our current overly cautious approach to measuring elements of the learning process that extend beyond content knowledge" (p. 1).

Noncognitive factors as composites. Isolating and measuring specific noncognitive factors is difficult. One reason is that precisely what is being measured is sometimes unclear (Messick, 1979; Conard 2006). Measurements of noncognitive factors report only manifest and often self-reported traits. For example, measuring self-evaluation and motivation (Täht & Must, 2010), does not account for the noncognitive reasons which may create the high or low scores, nor does it guard against myriad confounders. It does not distinguish whether the student is motivated or confident because she knows she is smart (cognitive), or because she just out-hustles everyone else (noncognitive). Kappe and van der Flier (2010) identify this as an inherent difficulty. "Weak and nonsupportive findings might have occurred because insensitive criterion measures were being used" (p. 142). They also conclude using the Big-Five (Costa & McCrae, 1992) personality traits as predictors of GPA is problematic and the traits should rather be used to predict specific performance in particular academic endeavors, such as team projects, skills training, and written work. This difficulty is unavoidable in discussions of noncognitive phenomena. It is difficult to separate cognitive from noncognitive factors, or to separate some noncognitive from other noncognitive phenomena.

While most measurements of noncognitive traits seem to be consolidations, there are also efforts to isolate and measure specific noncognitive constructs. Duckworth, Peterson, Matthews, and Kelly (2007) developed a measurement of grit, which they defined as "perseverance and passion for long-term goals" (p. 1087). This measurement of grit could predict achievement beyond measures of talent (Duckworth and Quinn, 2009). Conscientiousness is a particular noncognitive trait that is isolated and measured as a strong predictor of academic performance (Tross, Harper, Osher, & Kneidinger, 2000; Cheng & Ickes, 2009; O'Connor & Paunonen, 2007). However, conscientiousness could comprise multiple facets, including orderliness, self-control, dependability, impulse control, moralistic, and persistence, some of which are not accounted for in "an existing personality inventory" (Roberts, Chernyshenko, Stark, & Goldberg, 2005, p. 106).

The use of noncognitive is "already deeply embedded in educational policy circles, in the economics literature, and in broader discussions of student achievement. Though we agree with others' objections to this terminology, we feel compelled to use it" (Farrington et al., 2012, p. 2).

Farrington et al. admit that substituting a different word now "would likely confuse rather than illuminate our collective understanding of this important area of research" (p. 2). Surprisingly enough, different measures of noncognitive factors do correlate, indicating that there is a shared core construct (Sitzman & Ely, 2011). Part of the shared core construct of noncognitive factors is the ineluctable cognitive element. The composite nature of grades stems from the composite nature of noncognitive factors that influence them.

Grade Point Averages and Standardized Test Scores

Grades and standardized tests scores have been studied extensively as predictors of college success (Crede et al., 2010; Curie et al., 2012; deAngelis, 2003; Gaertner & McClarty, 2015; Meriac, 2012). Grades and standardized test scores are weighted differently from college to college, but they are prominent predictors of

> students' ability, typically measured in terms of SAT or ACT scores, and prior academic performance, typically assessed using high school GPA or high school graduation rank. Both variables have been shown to be independent, positive predictors of undergraduate grades … and there is no question about their utility in predicting academic success in college. (Harackiewicz et al., 2002, p. 562)

Standardized test scores and HSGPA remain among the most prominent and well-used metrics in college admissions (Balsa, Guiliano, & French, 2011; Gaertner & McClarty, 2015; Koenig et al., 2008; Schmitt et al., 2009). HSGPA and SAT together are better predictors than either alone (Kobrin et al., 2008). At least one study found that unweighted HSGPA was a better predictor of college GPA , than a weighted HSGPA (Warne, Nagaishi, Slade, Hermesmeyer, & Peck, 2014).

HSGPA and standardized test scores are the basis for the award of millions of dollars of scholarship funds for college-bound students annually (Volwerk & Tindal, 2012). For example, the NCAA (n .d.) requires that a Division I athlete must have a 2.00 HSGPA to receive athletic-based scholarship and a 2.30 college GPA to compete. A Division I athletic scholarship can provide a student with well over $100,000 in assistance. The President's scholarship at Concordia University of Nebraska (n.d.), worth $18,000 per year, is based on a combination of HSGPA and ACT/SAT scores.

High schools make use of widely varying weighting schemes. "While research has shown the statistical significance of high school grade point averages (HSGPAs) in predicting future academic outcomes, the systems with which HSGPAs are calculated vary drastically across schools" (Warne et al., 2014, p. 261). As an example of the different ways colleges might handle weighed and unweighted GPAs, Hutchinson Community College in Kansas uses weighted GPA only for scholarship purposes, and an unweighted 4.0 scale GPA for data collection and reporting. In a study which examined 232 of the largest 500 public school districts in the United States, Lang (2007) found that the majority of high schools implement some form of weighting system for grades or ranking, but that methods typically include inequitable premiums and inappropriate incentives for course selection. Lang reports that these flaws are well-documented, but largely ignored by researchers. Warne et al. (2014) conclude that comparison of HSGPAs from different schools was difficult if not impossible and that their "results demonstrated that unweighted HSGPAs were better predictors of CGPA" (Warne et al., p. 262). They recommend not using weighted HSGPAs when assessing the likelihood of student success in college.

Predictions are imperfect. According to Sparkman et al. (2012), HSGPA and ACT or SAT scores are "the best predictors of college success" (p. 644), but they account for only about 25% of a student's performance in college as measured by college GPA. Combining the HSGPA and a standardized test score improves the predictive strength of college GPA. HSGPA accounts for 19% of variance in college GPA, SAT accounts for 18% and together they account for 25% (Tross et al., 2000). Others have found different correlation levels. Sackett et al. (2009) conclude that cognitive tests including the ACTC and SATC are strongly correlated (r = .44) to college GPA. Kobrin et al. (2012) report correlations for first-year college grade point average (FYGPA) of r =.32 for SATC, and r = .46 for HSGPA and SATC together.

A problem with predicting college GPA based on HSGPA could be the structure of the research. Berry and Sackett (2009) report that the validity of these measures "has been underestimated because of previous studies' reliance on flawed performance indicators (i.e., college GPA) that are contaminated by the effects of individual differences in course choice" (p. 827). They controlled for this and found the percentage of variance 30% to 40% higher when predicting individual course grades. Their data set contained 5 million grades and 167,816 students. Their findings were that "SAT scores and high school GPAs together accounted for between 44 and 62% of the variance in college grades" (p. 822). This study shows the importance of how criteria are chosen and that composite scores, such as the composite college GPA, sometimes mask the predictive phenomena.

HSGPA and ACT/SAT Measure Different Things

One of the goals of standardized tests is to measure skill levels for individuals from different education environments under the same conditions, for example without undue influence from factors such as text book, method of instruction, approach, teacher, or curriculum

(College Board, 2010; Olson 2008). These same factors are an intrinsic part of grades. HSGPA and standardized tests have been shown to be independent predictors of undergraduate grades (Harackiewicz et al., 2002). Part of their independence stems from the amount of cognitive and noncognitive influence present in each measure. Grades and tests scores are correlated, but they do not measure the same things or rank students the same; there are several different dimensions on which they differ (Duckworth et al., 2012).

Gagné and St. Père (2002) found there is no relationship between IQ (cognitive) and motivation (noncognitive). "Grit, Duckworth discovered, is only faintly related to IQ. There are smart gritty people and dumb gritty people" (Tough, 2012, p. 75). Noncognitive measures and cognitive measures, such as IQ and test scores, are sometimes completely unrelated (motivation) or inversely or orthogonally related, for example, grit (Duckworth & Quinn, 2009) or intelligence and extraversion (Carter & Narramore, 1979). Steps taken to improve noncognitive factors may have little impact on cognitive results such as test scores (Holmlund & Silva, 2014).

Jaschik (2008) indicates that standardized tests are criticized because they do not emphasize noncognitive factors sufficiently. The Educational Testing Service (ETS), which produces the Graduate Record Exam, added The Personal Potential Index in 2009 to address this issue. Salisbury University became the first public university in Maryland to make the SAT optional, with university officials citing the belief that, while the SAT was a good predictor, high school grades were better. The SAT is optional for applicants with a 3.5 or higher HSGPA on a 4.0 scale (Dechter, 2007).

HSGPA more correlated with noncognitive factors. Farrington et al. (2012) wrote that course grades matter more than achievement test scores, this suggested that "grades do indeed capture something important about students that test scores do not" (p. 4). The review by

Farrington et al. (2012) examined noncognitive factors of "persistence, resilience, grit, goal-setting, help-seeking, cooperation, conscientiousness, self-efficacy, self-regulation, self-control, self-discipline, motivation, mindsets, effort, work habits, organization, homework completion, learning strategies, and study skills, among others" (p. 8). They found that HSGPA "is not only important in predicting whether a student will complete high school or college; it is also a primary differentiator by sex in educational accomplishment.

Duckworth and Seligman (2005) found that "when IQ and self-discipline were entered simultaneously in a multiple regression analysis, self-discipline accounted for more than twice as much variance in final HSGPA ($\beta = .65$, $p < .001$) as IQ did ($\beta = .25$, $p < .001$)" (p. 941). Cheng and Ickes (2009) found that students high conscientiousness and self-motivation, both noncognitive factors, had a higher HSGPA. Komarraju, Ramsey, and Rinella (2013) found that students had higher HSGPAs if they were more disciplined, determined, confident, and had better study skills. Conscientiousness, one of the Big Five traits, has been shown to have positive correlations with academic performance (Conard, 2006).

Standardized tests associated with ability or aptitude. Assertions about what the ACT and SAT exams measure are inconsistent. ACT claims their exam measures "achievement related to high school curricula, while the SAT measures general verbal and quantitative reasoning" (ACT, n.d.a, p. 1). They phrase this also as "what students are able to do with what they have learned in school, not abstract qualities such as intelligence or aptitude" (ACT, 2007, p. 1). Notwithstanding this, research shows a high correlation between ACT scores and IQ scores (Koenig et al., 2008). ACT emphasizes what their scores are useful for, namely "college readiness assessment" and predicting success in college (ACT, 2007). They eschew terms like *intelligence, aptitude, IQ, or reasoning.*

The College Board claims the SAT exams "test students' basic knowledge of subjects they have learned in the classroom—such as reading, writing, and math—in addition to how students think, solve problems, and communicate" (College Board, 2010, p. 1). However, according to Briggs (2009), "there is no explicit link made between the high school curriculum and the content of the SAT" (p. 10). The College Board has announced they are going to realign the SAT exam more closely with high school curriculum (Lewin, 2014, A1). According to the New York Times writer, "some changes will make the new SAT more like the ACT." The realignment could be because the SAT has lost ground to the ACT, which has sold more tests in the past two years. It could also a change in presentation, without substantial changes to the test itself.

Despite disclaimers, standardized test scores, including the ACT and SAT, are frequently referred to as measures of ability, as distinct from academic performance or achievement or what students have learned in the curriculum (Brown et al., 2008; Conard, 2006; NCES, 2011; Toomela, 2008). Standardized tests, including the ACT and SAT, are also referred to as "cognitive tests" and "aptitude tests" (Gagné & St Père, 2002; Harackiewicz et al., 2002; Komarraju et al., 2013). These tests are also referred to as cognitively loaded tests (Sackett et al., 2009). All parties agree that a primary use of the tests is to predict how well students will do in college (NCES, 2011; Maryland Higher Education Commission, 2007; Thomas et al., 2007).

Standardized tests highly correlated with IQ. A number of studies have found strong correlation between ACT and SAT exam scores and scores on IQ or other intelligence tests. "While the SAT and ACT were highly g-loaded, both tests generally predicted GPA after removing g. These results suggest that the SAT and ACT are strongly related to g, which is related to IQ and intelligence tests" (Coyle & Pillow, 2008, p. 719). Frey and Detterman (2004) found that SAT is "mainly a test of g" (p. 373), and they provide equations for converting SAT scores to estimated IQs. They found the correlation between SAT and the Armed Services Vocational Aptitude Battery (ASVAB) for 917 participants to be .82. Koenig et al. (2008) show a correlation of 0.61 between Raven's-derived IQ scores and ACTC, and a correlation of 0.77 between g-derived from the ASVAB and ACTC composite. They conclude that "ACT scores can be used to accurately predict IQ in the general population" (p. 153). This supports the assertion that ACT scores are a measure of native ability.

Internationally, Rindermann (2007) reported that results of international standardized student assessment exams, including the Programme for International Student Assessment (PISA) and Third International Mathematics and Science Study (TIMSS) exams are highly correlated with IQ. Kanazawa (2006) called the SAT "the preferred measure of general intelligence" (p. 594) because genuine IQ tests are not routinely given to representative groups in the United States. In sum, standardized test scores are a cognitive measure aimed at capturing student's ability.

Standardized tests highly correlated with each other. Briggs (2009) reported that the corresponding sections of the ACT and SAT are strongly correlated, between 0.8 and 0.9. Koenig et al. (2008) indicated that "a correlation of .92 was found between SAT I Verbal+Math and ACT composite scores in a sample of 103,525 students, and ACT Math correlated .89 with SAT I Math" (p. 153).

ACT and SAT scores do not change much. The ACT allows students to retake the ACT and reports that 57% of students who retake the ACT improve their composite score; 43% do not. For students who scored between 13 and 29, out of 36, the typical increase is about one point (ACT, n.d.b). They do not report whether subsequent retakes result in more increases. It should be noted that retaking the test means more money for ACT. For the SAT, the College Board (n.d.) reports similar results: 55% of scores improve, 35% drop, and 10% remain unchanged. They also report that higher scores are more likely to drop on retake and lower scores are more likely to improve and that retaking a second or third time the exam has diminished returns (College Board, 2010). There is a practical limit on how much any individual can improve their score.

The study by Briggs (2009) for the National Association for College Admission Counseling (NACAC) found that score increases for students subscribing to test preparation services were modest at best. The gains for the SAT were approximately 30 points (out of 1600 – without the writing portion). Briggs writes that most recent evidence shows, that for the ACT, private tutoring has a small effect of 0.4 points out of 36. According to Briggs, these improvements are too small to be distinguished from measurement error.

Test preparation services have little impact. Kaplan, the Princeton Review, Sylvan, the College Board, and ACT all provide test preparation courses or materials for the ACT and SAT. Costs for these programs can range from $30 for a book to $100+ per hour for coaching to $1,000+ for unlimited preparation. MacMillan (2010), writing for Bloomberg, estimates the annual test preparation market at $1 billion. Some services offer a guarantee if the student does not increase his or her scores, albeit by a relatively small amount. Those guarantees come in the form of additional free training, not refunds. Briggs (2009) shows there is no hard data to back up any claims of more than very modest improvements from test preparation services. Marte (2011) reports that test preparation services have significantly reduced their claims about improvements. Kaplan and the Princeton Review make no claims about specific increases, "calling that practice inherently misleading because it is difficult to collect accurate data" (Hechinger, 2009, p. 1).

This evidence supports the claim that standardized test scores are measures of a native and relatively immutable ability; and are not influenced by internal or external efforts to just work harder or perform better.

Changes in IQ scores. There is disagreement over whether intelligence changes over time or the reasons behind IQ score changes - if that happens (Cox, 2012). If ACT and SAT are highly correlated with IQ scores, then changes in IQ scores necessarily imply changes in SAT or ACT scores; likewise for lack of change.

Various researchers believe that cognitive ability is a "biologically determined characteristic of mind" (Toomela, 2008, p. 20). It is dependent on hormones (Fannon, Vidaver, & Marts, 2002), genetics (Trzaskowski, Shakeshaft, & Plomin, 2013), or combinations of these and other factors (Nyborg, 2007). Hormonal levels and individual events can change brain function, but genetic determinants do not change.

Heckman, Stixrud, and Urzua (2006) report that early childhood programs such as Headstart and the Perry Preschool Program do not improve IQ scores, however they have a positive impact on students from improved noncognitive skills, and thereby increase success in the students' social and economic lives. Jack Naglieri, research professor at University of Virginia, believes changes in IQ scores are possibly the result of changes in the way children think or function, rather than real changes in ability; he also makes a distinction between knowledge and ability (as cited in Cox, 2012). Moffitt, Caspi, Harkness, and Silva (1993) report that for the majority of children, changes in IQ are either small or due to unreliable measurements. Neisser et al. (1996) report the correlation between scores at the onset of adolescence age and the end of high school to be 0.96.

On the other side, research from Ramsden et al. (2011) shows that individuals can change their IQ scores and that these changes are associated with observable, physical changes in the brain. Rindermann (2007) reports that IQ generally goes up 3 points per year of school attendance. The Flynn effect posits that IQ scores for the general population increase slightly over time, approximately 3 points per decade (Flynn, 1999) . In these cases, when IQ scores do change, it appears the changes are gradual and the result of concerted effort or external factors.

Some of the disagreement on whether IQ scores can change turns on semantics or which particular "IQ" test is employed. Other disagreement stems from IQ scores potentially being an aggregate of different phenomena, for example, fluid intelligence or crystalized intelligence (Cattell, 1963). Fluid intelligence is the capacity to reason, identify patterns, use logic, and solve new problems. Crystallized intelligence is the ability to bring knowledge, particular skills, and experience to bear. This dichotomy has similarities to what the ACT and SAT allege to measure. As with cognitive and noncognitive factors, these phenomena do not exist independently. Crystallized intelligence improves with experience and new learned skills. Fluid intelligence may improve or diminish with age or neural function. Changes in either are presumably manifest in changes in IQ test scores.

Standardized test scores and IQ scores demonstrate a significant immunity from noncognitive factors. They do not change despite increased preparation or effort, test taking services, or early childhood programs. This is by contrast to grades, which respond well to programs such as early childhood education which impact noncognitive attitudes and behaviors.

Sex Differences in Cognitive and Noncognitive Factors

In many classrooms, the classroom climate, learning style, instructional style, and experiences offered to boys and girls may not address the needs of either sex. "This tunnel-vision view that all students learn in the same way regardless of gender, may be doing a disservice to our students" (Geist & King, 2005, p. 43).

Cognitive differences. The brain is the generally accepted locus of cognition. While direct observation that links the mind and the body (brain) is not yet possible, brain science is developing at a rapid pace and showing promise in understanding of sex-related cognitive phenomena.

Brain research. In the past, theories of learning and education were based on social, psychological, and philosophical perspectives of human beings. These frameworks did not have the benefit of close, internal examination of the brain and how it functions. It is not new information that "females tend to perform better on linguistic tests, including articulation speed, verbal fluency, grammar, and verbal production, while males tend to score higher on spatial tests, including mathematics, maze performance, and mental rotation" (Davidson, Cave, & Sellner, 2000, p. 510). What is new is the corresponding brain-based explanations. Diffusion tensor imaging, magnetoencephalography, functional MRI, and EEG all contribute to these discoveries (Ingalhalikara et al., 2014).

Brain structure. Brain research is also uncovering substantive differences in the ways that males and females receive, interpret, react to, and process information and the way the brain is structured, depending on sex (Gurian & Stevens, 2005). Nancy Forger, at UMASS Amherst, reported "at least 100 differences in male and female brains have been described so far" (p. 46). New technologies have produced large amounts of data showing differences in the male and female brain (Giedd, Raznahan, Mills, & Lenroot, 2012). Daphna (2011) notes "sex differences in the size of the brain and of specific brain regions, and in composition of neurons, neurotransmitter content, morphology of dendrites, number of receptors, etc." (p. 1). Females use different grey matter areas and more white matter for g-loaded tasks (Nyborg, 2007). In the typical female brain more areas are involved with linguistic function, there are more connections between the areas that handle emotion and language, and information is processed simultaneously in both halves of the brain, while men do this predominantly in the left side (Neu & Weinfeld, 2007).

Brain connections and processing. Ingalhalikara et al. (2014) have found not only is the structure, size, and composition of the brain different, but how various areas are connected is very different. Females have better connections between the two sides of the brain, which allows integration of various types of thinking, for example, intuition, memory, and multitasking. Males have more connections between the front and back and more intense activity within specific areas of the brain, allowing them to perceive and process complex information quickly and focus more intensely on a particular task. They found many of these wiring differences develop during adolescence when other secondary sexual characteristics emerge, influenced by hormones. Burton, Henninger, and Hafetz, (2005) report that some sex differences are prenatal in origin, including finger length ratios, which correlated with verbal fluency, spatial thinking, and SAT scores. Bull, Cleland, Mitchell (2013) report that sex differences exist in decision-making concerning numerical parity, magnitude, and estimations, and that this may be due to differences in the parietal lobes of males and females. Bell, Willson, Wilman, Dave, and Silverstone (2006) observed how and how much the different parts of the brain were activated during various cognitive functions. They report that activation levels in male brains were often more intense – even in carrying out a verbal fluency task. Their study "reinforces the fact that gender matching is essential in clinical functional imaging studies, and supports the idea of exploring male and female populations as distinct groups" (p. 537).

Critical thinking and memory. "From the cognitive scientist's point of view, … critical thinking [is] a subset of three types of thinking: reasoning, making judgments and decisions, and problem solving" (Willingham, 2007, p. 11). Willingham adds that critical thinking is highly dependent on memory, i.e., it is not possible to do critical thinking without content knowledge. Females seem more adept at remembering stimuli associated with a human or emotional content, as manifest in faces or facial expression; males are more adept at remembering objective information such as cars (Krohne & Hock, 2008; Dennett et al., 2012). Krohne and Hock (2008) showed that memory preserves connotation and context, and females remembered and associated aversive or ambiguous pictures with a negative connotation.

Interest and memory. There is some inconsistency in the research results in this area. Lovén, Herlitz, and Rehnman (2011) report that females are more likely to remember female faces (males are non-preferential here) and suggest this is due to greater perceptual expertise for female faces. Others have indicated that subjective interest in the object can be a differential and positive catalyst of memory (McKelvie, Standing, St. Jean, & Law, 1993). Dennet et al. (2012) accounted for and noted that perceived interest or expertise concerning cars was not the basis for better male recall; there was something in the object itself. They concluded that sex was the differentiator in the case of their study, not interest or expertise.

Sex as category or continuum. Although sex is often used as a binary category, there are issues with this classification. Bussey and Bandura (1999) frame the differences in the sexes as "a range of possibilities rather than … a fixed type of gender differentiation" (p. 676). Daphna

(2011) notes that brain science, which does underscore differences in male and female brains, does not make a clear bifurcation. It is more complicated than that. There are also similarities, variations within sex, and changes throughout life making the brain a moving target. Priest et al. (2012) note that this is not merely a brain-based issue. How and how much a student identifies with the paradigmatic social constructs which define gender can impact that students' noncognitive characteristics of motivation, interest, confidence, and self-image. "Children internalize gender norms and conform their behavior to those norms" (Hyde, 2014, p. 377).

The phenomenon of gender has a continuous element, albeit strongly bimodal, in the distribution of features of the brain and of social constructs. Structural changes in the brain occur during adolescence, at the same time that other physical secondary sex characteristics emerge and, as if that were not enough, social influences and gender-identify issues dominate a student's cognitive and noncognitive landscape. It is a wonder anyone makes it through middle school.

Implications in cognitive function. Brain discoveries have implications on how to present material so the receiver can process the information more effectively. If a group of students cannot learn as well with a cacophony of voices and simultaneous inputs, this results in frustration, which reduces motivation and produces dissatisfaction and failure.

The salient point for cognitive measurement is that while there are substantive mechanical differences, both sexes can learn mathematical processes, write papers, perform music, and execute scientific experiments, often with no noticeable difference in observable performance. It is hardly possible to know with any certainly whether a particular piano performance or a math exam was completed by a male or a female. The important feature for this study is that standardized test results for males and females are similar despite the differences in cognitive processing which produced them. Differences in cognitive mechanics do not necessarily result in significantly different results. One brain may process a question through path A, while another processes the same question through path B, but both brains arrive at the same answer; hence the measures of cognitive performance do not differ significantly. There is no research showing that males cannot remember faces, or that females cannot recall spatial characteristics of cars.

If there are palpable and nontrivial differences in cognitive function between the sexes, it is reasonable to examine any metrics, which measure cognitive functioning, separately for each sex. The index in this study relies on a measure of cognitive function. It should not be surprising if the predictive strengths of any subsequent analyses vary between the sexes.

Noncognitive differences. There are sex-based noncognitive differences that impact academic performance also.

Goal orientation and social-cognitive learning. Females and males often approach education with different goals in mind. Females are more likely to have mastery goals; males are more likely to seek performance goals (Bråten & Strømsø, 2008; Harackiewicz et al., 2002; You, 2010). Koul et al. (2012) report sex differences in goal orientation. While males have higher levels of mastery and performance goals in a few areas (e.g. physics), females have higher levels of mastery and performance goals than males in most areas (e.g. biology, reading). Performance avoidance was derived from a sense of competition and fear of failure. Chaput de Saintogne and Dunn (2001) found a similar result.

Darnon, Muller, Schrager, Pannuzzo, and Butera (2006) indicate that with epistemic conflict in a sociocognitive scenarios, mastery goals were more beneficial in resolving group disagreements than performance goals. Mastery goal orientation predicted an epistemic regulation of conflict, while performance goals predicted relational regulation of conflict. The epistemic manner takes into account recognizing other's views and others' competence; the relational manner emphasizes defending and asserting personal competence.

These results suggest that socio-cultural and social learning theories need to account for sex-related differences concerning individual and group work modes (Zumbar & Blume, 2008). Socio-cultural and social phenomena also have repercussions for levels of student engagement depending on classroom structure. This can impact grades and standardized test scores in areas which rely on assimilating information in that subject. When goal orientation, conflict resolution strategies, and performance avoidance are taken in concert, it helps account for a detrimental effect on males' academic engagement – when they perceive they are underperforming.

Self-efficacy. Boys often have higher self-efficacy beliefs, although these beliefs are not grounded in actual differences in ability (Bhanot & Jovanovic, 2009). Hyde (2014) reports that girls' math performance is equal to boys, but girls have lower math self-efficacy. Female students are more likely than males to attribute low marks to lack of ability or to task difficulty (McClure et al., 2011). Self-efficacy is perhaps the only noncognitive trait where boys exhibit an advantage over girls, although while boys have higher self-efficacy beliefs, these beliefs do not correspond to differences in ability or higher performance.

Self-efficacy and challenge. Challenge is important because it shapes peoples' decisions about whether to engage in challenging tasks (Hyde, 2014). Schmidt and Smith (2010) found as the challenge of the material in science class increases, girls become less engaged. Boys respond to both the perception of challenge and of difficulty of the material by intensifying their engagement. McGregor and Elliot (2002) echo that "men had stronger challenge construals than women" (p. 384). A construal is the manner in which a person perceives and interprets a situation.

Attribution. Sexes differ in how they identify and weigh the causes of individual success (McClure et al., 2011). Females tend to attribute success to more external factors, while males attribute success more to internal factors (Chaput de Saintogne & Dunn, 2001; Lloyd, Walsh, & Manizheh, 2005). For example, females are more likely than males to attribute both positive and negative scores to affiliation with the teacher (Lloyd, Walsh, & Manizheh, 2005).

Self-regulation. Self-regulation is a central noncognitive theme in academic performance. For Dash (2011) and Casillas et al. (2012), self-regulation is a metacognitive activity, a self-reflection about learning activities. It includes time management, mastery of learning methods, self-efficacy, and being goal-directed (Ruban and McCoach, 2005). It also includes specific behaviors like the ability to follow directions, work in groups, pay attention in class, and organize materials" (Jacob, 2002, p. 591). Self-regulation includes basic academic discipline, like going to class – which is more important than some might imagine. "Class attendance appears to be a better predictor of college grades than any other known predictor of college grades—including SAT scores, HSGPA, studying skills, and the amount of time spent studying (Crede et al., 2010).

Matthews, Ponitz, and Morrision (2009) report that self-regulation is playing an increasingly important role in predicting educational experiences and outcomes and leading to achievement. Dash (2011) concluded emphasis on student's cognitive reflection and self-regulation will lead to positive increases in achievement. Casillas et al. (2012) found that self-regulation added incremental validity in predicting grades. They distinguished between motivation and self-regulation, with self-regulation being the conscious personal monitoring and adjustment of the student's cognitive, behavioral, and emotional situation and progress. Schapiro and Livingston's (2000) study over four semesters and 300 students affirmed "the importance of dynamic self-regulation in attaining academic achievement" (p. 34).

There are sex differences in self-regulation with females demonstrating higher levels of self-regulated learning (Matthews, Ponitz, & Morrison, 2009; Zimmerman & Martinez-Pons, 1990). Wolters and Pintrich (1998) note that sex can explain differences in self-regulation in different subject areas. The research from Weis, Heikamp, and Trommsdorff (2013) showed that behavior regulation was higher for females than males. Cleveland (2011) notes that boys sometimes do not get an opportunity to develop emotional literacy, and this shortcoming impacts their interpersonal verbal communication and behavioral self-regulation, i.e., their ability to interpret and react appropriately to others' actions and words.

Research indicates that boys do enough, not more. Lackey, Lackey, Grady, and Davis (2003) report that the males with higher SAT scores show lower effort in maintaining course notebooks in an introductory engineering course. They postulate that higher prepared males view spending time on a notebook that has a low point value in the course grading scheme as unimportant for them to achieve the grade they desire.

Culture and negative self-regulation. Culture influences the development of self-regulation and sexes have distinct difference in cultures (Connell & Messerschmidt, 2005; Weis et al., 2013). "Men and women interact differently with the learning environment. Women's standards and goals are responsive to social and environmental influences. Men seem relatively indifferent but check their performance against strongly internalized standards" (Chaput de Saintogne and Dunn, 2001, p. 1024). From a different perspective, what is termed "low" self-regulation for males could actually represent a high degree of self-regulation – in a negative academic sense; not the same as low self-regulation. There is lack of positive example in the media and culture to inform boys about what it means to be masculine and which can positively impact their self-image, development, and identity (Fitzclarenge, 1999). There is a sense in which male culture actively and negatively impacts performance (Van Houtte, 2004; Hickey, 2008). Boys check their performance versus an internal standard, however, they also respond to male socio-cultural influences, such as hegemonic masculinity.

Hegemonic masculinity describes the set of regulative practices that supervene on a group of males, informing the members' personal senses of identity, including which behaviors are expected and which are not tolerated (Juelskjær, 2008). "It is a regulative ideal influencing both how boys think they must act to be acceptable as boys, as well as how concrete practices are constituted and negotiated in social interaction" (p. 53). Boys noncognitive behaviors are often influenced by peer pressure, examples from parents and media, and the reactions which their behavior causes (Skelton, 1997). Extensive evidence has accumulated showing how schools, sports, and other cultural milieus assist in creating and perpetuating hegemonic masculinity and its negative impact on boys (Connell, 1995; Hickey, 2008).

Cleveland (2011) refers to this set of hegemonic rules as "the Code." Neu and Weinfeld (2007) capture this code in specific edicts such as do not cry, do not ask for help, and do not reach for reassurance. "Unfortunately, thanks to the Code’s demands for emotional stoicism …many boys never get the chance to develop emotional literacy, and its absence affects the productivity of their interpersonal verbal communication as well as behavioral self-regulation" (Cleveland, 2011, p. 137). Van de gaer et al. (2006) report that boys' underachievement in language is related to negative school-related attitudes. "Especially, the more intelligent boys in the lower tracks appeared to be demotivated" (p. 307).

Combining noncognitive factors. High self-efficacy is positively related to performance, but it can be a double-edged sword if it conflicts with actual success, or lack thereof. "Since men are concerned with maintaining an appearance of competence, the suggestion of performance failure is especially threatening" (Chaput de Saintogne & Dunn, 2001, p. 1029). When boys' high self-efficacy, competitive instinct, fear of failure, and internalized attributions are confronted with lack of success and witness of girls' higher relative performance, a coping mechanism can emerge to protect the male psyche (Cramer, 2006). This phenomenon is commonly called "sour grapes." This escape relieves the cognitive dissonance in the situation, but causes boys to devalue and be less engaged in education. This is reinforced in the Code (Cleveland, 2011).

Content influences interest and engagement. An important factor of success in education is interest in a topic or discipline (Harackiewicz et al., 2002). According to British Education Secretary, David Blunkett, "instead of trying to interest boisterous teenage boys in the novels of Jane Austen, they should introduce them to … tales like Robert Louis Stevenson's Treasure Island and Sir Arthur Conan Doyle" (Birmingham Post, 1999). Ainley, Hillman, and Hidi (2002) found sex as the factor most closely related to topic interest and that boys' long and sort term recall was better than girls when the main character of a story was a violent male. This phenomenon develops very early in males. Hanson and Zambo (2010) report that, for preschool boys, books containing positive male archetypes elicited almost twice as many meaningful word utterances as androgynous characters.

Daniels et al. (2008) found that sex was a non-significant covariate in all analyses – except boredom. In his article about the masculine propensity to resist schooling, Juelskjær (2008) indicates boys do not lack potential, instead schools need to activate that potential.

Chaput de Saintogne and Dunn (2001) write that achievement in males demands arousal and a congenial school atmosphere may not be the most conducive to this. "It seems that negative events that are persistent and pervasive are able to cause arousal through a fear of criticism that results in high achievement behavior… This suggests that a generally supportive learning environment may lack the threat and arousal necessary for men to achieve in our exams" (p. 1029).

Male underperformance in education. Matthews et al. (2009) write that "research today on gender and education in kindergarten through 12th grade school settings reveals that girls tend to build stronger relationships with teachers, attain higher grades, achieve higher levels of education, and progress better scholastically overall than boys" (p. 689). What follows here are some specific areas in which performance might be measured and where genders are, or are not, at similar levels.

Cognitive and noncognitive measures. There are non-trivial cognitive differences between sexes, for example, brain-based differences in the way that language is processed. However, these kinds of differences do not render any conclusions about performance. Reiss, Abrams, Singer, Ross, and Denckla (1996) note that total cerebral volume is 10% greater in boys than girls, that cortical grey matter is the biggest contributor to this difference, and that IQ is correlated with total cerebral volume, in particular with cortical grey matter. This seems to favor males, but no ascription of "performance" is tied to brain capacity or IQ or brain functioning. Performance does not turn on these factors and neither males nor females have a performance advantage based on cognitive factors. However, noncognitive factors are different matter. As reported in the previous section, numerous noncognitive factors of performance including self-regulation, paying attention in class, affiliation with the teacher, and not resisting education seem to favor females.

Standardized tests. Standardized test scores revealed modest differences between sexes. For example, the 2013 composite ACT scores for males and females were identical at 20.9 (ACT, 2013c).

The SAT. 1,660,047 graduating seniors took the SAT in 2013 (SAT, 2013b). 53% of these were female, 47% male. The scores are as follows:

Table 1

Comparison of SAT Scores for Males and Females

		Critical Reading		**Math**		**Writing**		**Composite**
	Total	Mean	SD	Mean	SD	Mean	SD	
Male	776,092	499	117	531	121	482	115	1512
Female	883,955	494	112	499	114	493	112	1486

Note. This data is from the SAT® Total Group Profile Report. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/2013/TotalGroup-2013.pdf

The Critical Reading (d = .044) and Writing (d = -.097) scores are similar. The Math scores are somewhat different (d = .273).

The ACT. 1,799,243 graduating seniors took the ACT in 2013 (ACT, 2013c). For the first time ever, males did not outscore females in the composite score, with the mean composite score for both sexes at 20.9.

Table 2

Comparison of ACT Scores for Males and Females

	Total	English	Math	Reading	Science	Composite
Male	835,431	19.8	21.4	20.9	21.2	20.9
Female	954,919	20.6	20.5	21.4	20.4	20.9

Note. This data is from the ACT Profile Report – National. Retrieved from http://www.act.org/newsroom/data/2013/pdf/profile/National2013.pdf

The ACT did not show Standard Deviations for scores. However, Cohen's *d* for the Composite ACT for males and females is obviously zero.

National Assessment of Educational Progress (NAEP). The average NAEP scores for 17 year-old males are 308 and 283, with standard errors of 1.0 and .8, respectively (NCES, 2012a). For females the corresponding scores are 283 and 291, with standard errors of 1.1 and 1.0, respectively. The standard errors are quite small as the sample size is large. These scores do not show a significant practical sex difference.

Grades. The NCES online statistical generator, QuickStats, generated the report in Table 3. It shows that, for 2009, the latest year available from QuickStats, the proportion of females in the highest grade category was 27% higher than the proportion of males in that category. Fortin, Oreopoulos, and Phipps (2013) report a "growing gender disparity in high grades in high school" (p. 2).

Table 3

High School Grade Point Average by Sex as recorded by NCES

HS GPA Estimates	0.5-0.9	1.0-1.4	1.5-1.9	2.0-2.4	2.5-2.9	3.0-3.4	3.5-4.0	Total
Male	0.2	0.9	3.3	15.5	16.0	35.4	28.7	100%
Female	0.1	1.1	3.0	10.5	13.5	35.2	36.5	100%
Both Sexes	0.2	1.0	3.1	12.7	14.6	35.3	33.0	100%

Note. Rows may not add up to 100% due to rounding. The computations are from the NCES Quickstats on 3/29/2014 at http://nces.ed.gov/datalab/quickstats/

Summary

The research in education, biology, sociology, and psychology shows substantive differences for males and females; some pre-existing the formal education years and some extending past that. Because of this, examining measures of performance disaggregated by sex may improve understanding and insight. It is reasonable to suspect sex is a differentiator in any measure of performance that is based on noncognitive factors.

The literature indicates that both grades and standardized test scores are useful in predicting college performance. Standardized test scores, such as the SATC, are shown to be highly correlated with cognitive measures such as intelligence, as measured by IQ scores and other indices. Precisely what IQ and SATC tests measure is subject to debate, but they are often referred to as tests of ability, aptitude, and cognitive capacity. These tests are subject to little influence by noncognitive factors. On the other hand, HSGPA is a measure of performance and is highly influenced by noncognitive factors.

CHAPTER THREE: METHODOLOGY

Introduction

The purpose of this non-experimental, correlational study was to predict success in college by creating and examining a new metric from existing and commonly used measures. The new metric was an index termed the Individual Performance Index (IPI). This was the sole predictor variable. The detailed mechanism for creating this index is described in Instrumentation; basically it was constructed by taking the quotient of the student's HSGPA, as a percentile within the incoming freshman class, and the student's composite SAT (SATC) score, as a percentile within the incoming freshman class. The criterion variable was the first-semester college cumulative grade point average (FSGPA).

Design

This was a non-experimental, correlational study. Archival data was used to create an Individual Performance Index (IPI) using each student's cumulative high school grade point average (HSGPA) and composite SAT (SATC). The null hypothesis that there was no significant predictive relationship between the IPI and first-semester college GPA (FSGPA) was then tested (Gall, Gall, & Borg, 2007). Subsequently, the same predictive model was tested separately for males and for females. Standardized test scores and HSGPA are the most widely used and relied upon indicators that a student is prepared college (Gaertner & McClarty, 2015). Correlational designs are often used when determining the predictive validity of HSGPA and SAT scores for college grade point averages (Bridgeman, Pollack, & Burton, 2008; Higdem, et al., 2016; Noble & Sawyer, 2004; Warne et al., 2014). Thus, this design was an appropriate choice for the study.

Threats to validity for correlational studies include errors-in-variables, omitted variables, and simultaneous causality bias (Bascle, 2008; Feng & Xue, 2014; Warner 2013). Errors-in-variables arises from errors in the measurement of predictor variables. This threat was minimized here because the IPI is constructed from HSGPA and SATC scores. Both of these are valid measures and predictors of first-semester college grade point average (FSGPA) (Gaertner & McClarty, 2015). The reliability of the SATC score ranges from 0.89 to 0.92 (College Board, 2013). The College Board asserts that a "student's SAT score, combined with his or her high school record … is the best indicator of how well that student will do in college" (College Board, 2010, p. 1). The significant correlation between SAT test scores and IQ test scores is also acknowledged (Coyle & Pillow, 2008; Frey & Detterman, 2004).

The potential for omitted variables was not trivial because HSGPA is influenced by a number of noncognitive factors (Farrington et al., 2012; Gaertner & McClarty, 2015; Lipnevich & Roberts, 2012). The predictive model in this study only examined the index as a predictor in order to establish its baseline relevance.

Simultaneous causality was not a factor since the predictor variable was developed chronologically prior to the criterion variable. However, sample selection bias might have been another a threat to validity. Sample selection bias occurs when some values of the criterion variable are missing because of the process of sampling (Certo, Busenbark, Woo, & Semadeni, 2016). Sampling in this case was random and all non-defective data records (without missing or invalid data) had equal probability of selection. The sample selection bias in this study was restricted to the selection bias in the admissions decisions and matriculation of the particular incoming class.

Indices

The first procedure in this study was to create an index number for each student. This index was derived from the student's HSGPA and SATC, but it is a new metric which cannot be constructed from a linear combination of either or both of these scores. Some background on the ubiquity, usefulness, and power of indices is appropriate.

> If you should say to a mathematical statistician that you have discovered that linear multiple regression analysis and the analysis of variance (and covariance) are identical systems, he would mutter something like, "Of course—general linear model," and you might have trouble maintaining his attention. If you should say this to a typical psychologist, you would be met with incredulity, or worse. (Cohen, 1968, p. 426)

Cohen is saying that what is a commonplace in one discipline might seem out of place in another. The use of indices is found in many places in education as well as myriad other disciplines.

The index created in this project was reasonably straightforward and has a basis in the literature concerning cognitive and noncognitive measures of academic performance. This project examined whether the index was a useful predictor of college performance for incoming freshmen. The lack of a specific interpretation of the index does not invalidate it, nor impact its utility. That shortcoming simply puts the index in the same category as grade point averages, standardized tests scores, and many other indices enumerated here, many of which are subject to debate and disagreement.

The terms "index" and "ratio" are frequently used interchangeably and sometimes consecutively. "The terms Gini Index, Gini ratio, and Gini coefficient are equivalent and will be used interchangeably throughout this article" (Wan, 2001, p. 361). Brannian, Schmidt, Kreger, and Hansen (2001) use the term "body mass index ratio" in the title of their journal article (p. 1819). A ratio is simply one number divided by another; a quotient. The resulting number provides the index of interest. Sometimes an index is not called an index at all, as in the case of effect sizes, z-scores, and percentages – all of which are indices.

Uses and Examples of Indices

Indices are designed to make numbers and comparisons of numbers more meaningful by make them relevant to some benchmark. Indices account for things like scale, context, and other artifacts that make direct comparison of numbers misleading. Linear correlation can be viewed as an index. "The measure *r* can then be interpreted as an index of linear association between X and Y" (Kleinbaum, Kupper, Nizam, & Muller, 2008, p. 92).

Apply a scale. Interpreting a measurement is difficult without knowledge of scale. For example, an ACT Composite score of 32 means little without knowing the maximum score is 36, the average is 21.7, and a score of 32 is in the 98th percentile (ACT, 2013b). Even percentages are misleading if they are calculated linearly in the context of a different model. Here the linear percent of 32 out of 36 is the 89^{th} percentile, but the model is actually Gaussian (normally distributed), not linear, and the score is in the 98th percentile.

Another example of this is effect size, for example, Cohen's *d* (Cohen, 1988). Two samples with means that differ by 15 could infer the two sets are comparable or very different; it depends on whether the pooled standard deviation is 7 or 140. The difference of means is misleading until it is indexed against a measure that adds relative scale. The interpretation of a score that is 25 points above the mean is different if it is indexed, that is, *z*-scored, with a standard deviation of 9 or a standard deviation of 126.

A company's profit or loss of $1 million is not meaningful until it is compared with some measure of the company's size. Hence, the common business indices of Return on Assets (ROA) and Return on Investment (ROI). A quarterly profit of $1 million would be a disaster for General Mills, but a spectacular result for Sam's Bait Shoppe.

Put numbers in context. The theoretical basis for an IQ score is the quotient of a person's mental age divided by their chronological age (Neisser et al., 1996). This index is called the Intelligence Quotient, but it could also be called an index or ratio. The IQ score shows how adding context is useful. Any raw measure of a student's vocabulary would be more meaningful with information about the student's age or native language. For one athlete, running a 15 second 100 meter dash is exceptional and probably wind-assisted, for another it is a poor showing. It all depends on that athletes ability, which is the context.

Make comparisons between raw numbers. This use of indices is a two–step process. First, a raw number is converted to an index, then it is compared to another number, similarly converted. Company and industry financial ratios are created in business and economics from raw numbers, for example the quick ratio, debt to assets, return on investment, return on assets, and earnings per share (Financial Ratios, 2011). These indices allow for comparison between companies, between industries, between companies and their industry, and within a company over time.

Remove artifacts from numbers. Some indices help account for artifacts in raw numbers to improve comparisons and interpretations. The Consumer Price Index (CPI) is used to compare prices from different years. The U.S. Bureau of Labor Statistics (BLS, n.d.) actually publishes thousands of CPI indices each month and they are used for critical purposes such as adjusting Social Security payments, federal retirement benefits, and Treasury Inflation-Protected

Securities returns. The BLS uses the same methods as other Organization for Economic Co-operation and Development and European Union nations to compute the American CPI (BLS, n.d.). According to the BLS, the commonly used Consumer Price Index for All Urban Consumers (CPI-U) for 2012, using 1982 as the base year (=100.000), is 229.594 (BLS, 2014). The actual value of something that cost $100.00 in 1982 and $229.59 in 2012 did not change. The nominal price of a metric ton of corn in October 1981 was $111.41 (McMahon, 2011). It was $274.85 in October 2011. This is a price increase of 146.7%. But, when these prices are adjusted to 2011 dollars by removing inflation, the adjusted prices are $270.08 and $274.85, respectively, which is an increase of only 1.7%. Without removing inflation, the price increase appears to be 86 times higher than it actually was.

Track changes. Indices are used for longitudinal comparison to show increases and decreases over time. The United Nations Development Programme (2011) has developed the Human Development Index and the Gender-related Development Index to create "a simple composite measure that includes health, schooling and income" (p. 23). This national index allows United Nations' monitoring agencies to track changes over time in these targeted areas of concern.

Aggregate numbers. Indices are frequently used to aggregate numbers or represent information derived from various factors. "The power of using indices as management tools clearly resides in their ability to capture the information contained in a large number of variables in one number" (U.S. Department of Energy, 1995, p. 1-59). The CPI is a prime example of this, as thousands of numbers are collected to compute a single CPI. Standardized test scores and grades are other examples. They are influenced by myriad cognitive and noncognitive factors, including some that are not explicitly identified. "More typically, an index is complicated because it combines a number of different types of indicators" (Covalent, n.d., p.3).

Quantify phenomena that are essentially qualitative. Enterprise Worldwide, an international association of accountants and advisors, reports that Return on Employees is a "new concept that is gaining attention in many companies" (Gillum, 2012). This index accounts for hard costs of employees and difficult to measure things like replacement cost, costs to develop the employee, and even the employee's impact on morale. Figge, Hahn, and Schaltegger (2001) have reported Return on Employees and Return on Government for companies like Volkswagen, Daimler-Chrysler, BMW, and Porsche on behalf of the United Nations Center for Stability Management. The Graduate Record Exam and Educational Testing Services have agreed to start using the Personal Potential Index (Jaschik, 2008). Other examples of qualitative phenomena being indexed are the Human Development Index and the Gender-related Development, both from the United Nations, as well as, the Work Problems Index, Marital Conflict Index, and Depression Index (Kleinbaum et al., 2008).

Predict and prevent problems. The Work Ability Index (WAI) is a multidimensional and multidisciplinary index of work ability which combines seven different factors, each measured on a different scale (Ilmarinen, 2007). It is used to predict work disability, retirement, and mortality, and to maintain, reestablish, and promote the work ability of employees (Ilmarinen, 2007; Hasselhorn, Müller, Freude, Tempel, & Kaluza, 2005). It was developed as a tool to promote working ability during aging, which has become a more important concern due to changes in globalization and working lives being extended (Ilmarinen, Tuomi, & Seitsamo, 2005). It can be used to prevent impairments for individuals as they progress through their working lives (Nunes, Costa, & Puga-Leal, 2011).

Create standards for social policy. The United Nations monitors compliance with its social policies and programs by tracking the Human Development Index and the Gender-related Development Index. They use the basis of equitable or reasonable-to-expect percentages in creating these. For example, if a group represents 25% of the population, they should represent 25% of the income, 25% of the educational aid, and 25% of the political power. The Gini ratio is a similar social index for monitoring regional income inequalities in China (Wan, 2001). It has a range of 0 to 1, with 1 being perfect inequality. "Prominent economists in China seem to take a Gina value of 0.2 as the critical point…a value smaller than 0.2 signals reasonable equity and a larger value is used in appeals for policy action" (p. 377).

Issues with Indices

Indices are robust tools for enhancing meaning and usefulness of measurements and statistics, but they are not without their own complications.

The meaning of an index may not be precise. The Washington Post reports, "Arbitrary stock index hits arbitrary number. But it's not as meaningless as you might think" (Irwin, 2013). Every index has a specific method for its computation, but what the index stands for may not be precisely defined, for example, the Global Peace Index, from the Institute for Economics and Peace. There are thousands of CPIs computed each month and each can involve more than a thousand numbers. In addition, the bases for calculating each CPI, including what numbers are used or not, is subject to change (BLS, n.d.). The BLS maintains that the best measure of inflation will vary depending on usage; it is not always the same CPI.

It is difficult to say precisely what the SAT and HSGPA measure, or do not measure, but they are indisputably common metrics for college-bound students. Not having precise explanations of indices does not preclude their usage. For example, the comparative index of educational performance created by Pearson and the Economist Intelligence Unit "is anything but a straightforward exercise" (Pearson, 2012, p. 6).

Disagreement on how to construct. Dijkstra (2002) writes that the construction of the United Nations Gender-related Development Index (GDI) has weaknesses and proposes that a new measure, the Standardized Index of Gender Equality (SIGE), can draw "on the good aspects of GDI and GEM [Gender Empowerment Measure] while at the same time attempting to avoid their methodological limitations" (p. 301). The Work Ability Index (WAI) has been shown to be valid and reliable, but the dimensionality of the WAI is subject to disagreement (Martus, Jakob, Rose, Seibt, & Freude, 2010). The Body Mass Index (BMI) has widespread international usage, yet it ignores muscle mass, body type, and misclassifies nearly half of women and 20 percent of men as obese (Shah & Braverman, 2012).

Summary of Indices

Index numbers have been around a long time (Fisher, 1923). They are created in countless ways ranging from simple to complex, and measure everything from commodities to subjective human phenomena. They are used to measure simple physical quantities and also measurement-resistant notions like intelligence. They are used in nearly every discipline, including education, economics, business, psychology, sports, and politics.

The primary purposes of indices are to provide context for numbers, simplify and quantify phenomena, and to improve comparisons, predictions, and monitoring. They make numbers relative so they can be more relevant and more meaningful. They are limited only by imagination, and driven by the need to quantify an often qualitative world. The use of indexing in this study is reasonably straightforward. It is simply anchoring one measure of general academic performance to a measure of individual ability in order to see if more profitable comparisons can be made.

Research Questions

The research questions examined the predictive validity of the Individual Performance Index (IPI) and first-semester grade point average (FSGPA). This was done for all students and then separately for males and females.

RQ$_{1:}$ Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average?

RQ$_{2:}$ Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males?

RQ$_{3:}$ Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females?

Null Hypotheses

The null hypothesis for this study are:

H$_{01}$**:** There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average.

H$_{02:}$ There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males.

H$_{03:}$ There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females.

Participants

The sampling frame for this study was a random sample taken from a freshmen cohort of 544 students attending Northeast College (a pseudonym). This convenience sample of students was from a mostly middle class socioeconomic status and recent high school graduates. The ages of the freshmen ranged from 17 to 22, with 96.7% of them (n = 526) between 18 and 19 years old. Hispanic students accounted for 5.1% of the population (n = 28), while the total of all Black, African American, Asian, American Native, Hawaiian or Pacific Islander accounted for 11.4% (n = 62). Females made up 56.4% of the cohort (n = 307); males made up 43.6% (n = 237). The demographic information collected included ethnicity (Hispanic or non-Hispanic), race (White, Black or African American, Asian, American Native, Hawaiian or Pacific Islander), geographic region (assigned by College Board). The data collected from this sample was archival.

Sample Size

There are varying opinions on the appropriate sample size for a bivariate correlation analysis (Field, 2009). There are tradeoffs between sample size, level of significance, directionality, and effect size. These factors are mathematically related and any three of them will determine the other (Gall, Gall, & Borg, 2007). For this study, with an α of .05, medium effect (.7), and statistical power of .7, the sample size should be set to 66 (Gall, Gall, & Borg, 2007, p. 145). However, Warner (2013) recommends sample sizes of at least 100 where correlations are reported. The selected sample size of 100 for each analysis satisfied these guidelines. Also, a sample size of 100 falls between one quarter and one third of the population after cases were eliminated for missing or invalid data.

Setting

The participants in this study attend Northeast College (a pseudonym), which is a 4-year private college that offers over 30 majors based on a liberal studies curriculum. The first semester curriculum for all majors is similar and liberal arts focused. The vast majority of faculty have doctorates, class sizes average less than 20, and most students live on campus. This is a northeastern regional college that draws students primarily from its own and contiguous states. The college is categorized as having "moderately difficult" admission standards according to Peterson's Publishing. The Carnegie enrollment profile for this college is classified as "Very High Undergraduate." Carnegie lists 700 United States colleges in this category, including the study institution.

This college was selected because of its homogeneous population. The incoming cohort is predominantly white, non-Hispanic, close in age, from similar Socio-Economic Strata, and, in many cases, from the same or similar towns, high schools, and cultures. This is significant as the index created in this study is influenced by various cognitive and noncognitive factors (Duckworth, Quinn, & Tsukayama, 2012, Farrington et al., 2012; Komarraju et al., 2013), which are inherently intertwined with demographics. Similar demographics helped to reduce confounders.

Instrumentation

Data was collected from a designated representative of the college. It included no personally identifiable information. The data elements collected included sex, month and year of birth, high school cumulative GPA, composite SAT, high school class rank, fall semester GPA, spring semester GPA, first year cumulative GPA, ethnicity (Hispanic or non-Hispanic), race

(White, Black or African American, Asian, American Native, Hawaiian or Pacific Islander), and geographic region (assigned by College Board).

Criterion Variable: First-semester College GPA

The criterion variable was the first-semester grade point average (FSGPA). This was recorded in the data supplied by the study institution on a 4.0 scale.

Predictor Variable: Individual Performance Index (IPI)

The single predictor variable for each of the hypotheses was the Individual Performance Index (IPI). The index was constructed by taking the quotient of the percentile rank of each student's HSGPA divided by the percentile rank of that student's SATC score. This requires four steps in SPSS. The first step was to compute the mean and standard deviation for both the HSGPA and SATC for the entire freshman cohort. Next, the HSGPA and SATC raw scores were converted into z-scores by subtracting the mean from each and dividing that result by the standard deviation. Then, the built-in SPSS function "CDFNORM(z-score)" was used to convert the z-scores into percentiles. This function calculates the probability that a variable with mean of 0 and standard deviation of 1 would less than the z-score entered. The value returned by this function, multiplied by 100, is the percentile of that z-score entered.

As an example of this calculation, assume the mean of the SATC scores is 1700 and the standard deviation is 150. A SATC score of 1820 would translate to a z-score of +0.80.

$$z - score = \frac{1820 - 1700}{150} = 0.80$$

Approximately 57.63% of the normal distribution is below a z-score of +0.80, therefore the CDFNORM(0.80) would return a value of 0.5763. Multiplying 0.5763 by 100 yields a percentile of 57.63%. The same system was used to convert HSGPA from a raw score into a percentile.

These calculations were based on the study population, rather than national percentiles, because this group may not conform to the national distribution's mean and variance. The population in this study was not projected as representative of nationwide phenomena. After the percentiles were calculated for the student's HSGPA and SATC scores, the IPI was created by dividing the percentile rank for the HSGPA by the percentile rank for the SATC.

For example, if a student was 73rd percentile for HSGPA and 73rd percentile for SATC, then the index (IPI) will be 1.0 because 73/73 = 1.0. This could be viewed as performing as expected because the student's HSGPA position among the other students was aligned with the student's SATC position among the same students. If a student was 50th percentile for SATC but 75th percentile for HSGPA, then the IPI would be 75/50 = 1.5. This student was only 50th percentile in ability, but 75th percentile in high school grades, indicating overperforming compared to what might be expected. This method of constructing the IPI has an advantage that percentiles are commonly used and easily understood position rankings.

As an example from the data, there was a student who was at the 98.4th percentile for HSGPA, and at the 98.2nd percentile for SATC score. A very high ability student performing at the top of the group, as expected. There was also a student who was at the 27.4 percentile HSGPA and 27.3 percentile SATC. This was a student with very different ability and performance than the previous student, but whose performance was aligned with his or her ability. The IPI scores for these students were 1.00 and 1.00, respectively.

Control Variables

The control variable utilized in Research Questions 2 and 3 was sex. This divided the population into two categories, male and female. Sex was self-reported as part of the institution's application process.

Procedures

The researcher contacted the registrar of Northeast College to ascertain preliminary willingness to participate in the study. The researcher received formal approval from the Northeast College IRB (see Appendix A) and the Liberty University IRB (see Appendix B). After obtaining IRB permissions, the researcher obtained the raw data from the CIO at Northeast College. The data was provided in a PC-compatible electronic file format. After receiving the data, the researcher sampled the data according to the sampling method described above. Analysis of the data proceeded.

Data Analysis

The primary vehicle for data analysis was SPSS (IBM SPSS, Version 23). Some other tasks, including initial screening of the data set for incomplete data was performed with Microsoft Excel 2010. Data records with incomplete or invalid data elements were discarded.

High School Grades – Weighted vs. Unweighted

High schools have myriad computational schemes for computing the grade point average, including various schemes that weight some grades differently than others. According to the registrar of Northeast College, 100% of the HSGPA figures in the data base are taken from official high school transcripts rather than self-reporting. If there was a weighted HSGPA, then that figure was recorded, otherwise the unweighted figure was recorded. However, no record was kept of whether the HSGPA was weighted or unweighted. Note that a HSGPA of 3.8 could be weighted or unweighted. Even if there is a weighting scheme, some students might not have taken weighted courses and their HSGPA will be the same as if no weighting scheme existed. The Northeast College registrar indicated that specific coursework from transcripts was examined and that college prep course grades carried more weight, albeit heuristically, if not quantitatively. This aligns with the 2014 report from the Director of Research for the National Association of College Admissions Counseling: "The top factors in the admission decision were (in order): grades in college preparatory courses, strength of curriculum, standardized admission test scores, and overall high school grade point average" (Clinedinst, 2014, p. 3). Clinedinst (2014) also reported that more than 80% of post-secondary schools rated college preparatory course as significant. It is important to note this study is not a referendum on the vagaries of calculating grade point averages. It is simply to ascertain whether the reported and indexed HSGPA can serve as a significant predictor for college grades. For this study, the HSGPA recorded in the data base at Northeast College is the statistic used.

Recoding High School Grades

The grades recorded in the data base for Northeast College represented a range of reporting schemes, although most of them were on a 4.0-5.0 scale. Some of the HSGPAs in the Northeast College data base were based on a 100-point scale. It is reasonable to impute values for data when an equivalent score can be computed or closely estimated (Warner, 2013). In this situation, the 100-point scores were converted to a 4.0-5.0 scale. The formula used to convert from a 100-point scale was:

$$\text{4-5pointGPA} = (\text{100pointGPA} - 55) \times 0.10$$

This formula converted a 75 on the 100-point scale to a 2.0. It converted an 85 on the 100-point scale to a 3.0, and a 95 on the 100-point scale to a 4.0. This translation followed the conversion table from the College Board at http://www.collegeboard.com/html/academicTracker-howtoconvert.html with the exception that it produced a continuous transformation rather than a stepwise one. It is also aligned with Northeast College's grading scheme which assigns numerical values from 0.0 to 4.0 for letter grades A – F, with pluses and minuses, excluding A+

and D–. A continuous function was used because the 4.0-5.0 grade averages in the data base formed a more or less continuous distribution. Because the Northeast College data base allows 4.0-5.0 scale scores above 4.0, scores above 95 were not capped at 4.0. However, there was one data point with a 100-point scale HSGPA of 108.4. With the previous formula, this calculated to a 5.34 GPA, which seemed excessive as it was the only score above 5.0. It was recoded to 4.93, which is its corresponding percentage on a 110-point weighted scale. This was still the highest GPA in the file. This imputation also kept it within 3.0 standard deviations of the mean so that it would be used. There were no other compelling reasons to discard it completely.

Assumption Testing

Prior to conducting the bivariate correlations, assumption testing was conducted. The assumptions to be made in Pearson correlational analyses are normality, linearity, homoscedasticity, and lack of outliers. Each of these assumptions was examined for tenability with results reported in Chapter 4. Normality means the population distributions for the predictors and the criterion are normal or bell-shaped. Tests for normality include the Shapiro-Wilk and Kolmogorov-Smirnov tests. Histograms are also examined to review the normality assumption. Linearity means the change in one variable is a constant multiplied by the change in the other variable. Graphically, this means that a "best fit" line drawn through the middle of a scatterplot of two variables will appear to be a straight line, rather than curved. Homoscedasticity means that the variance around the line of best fit is consistent. The Pearson correlation analysis assumes that the distance of the data points from the best fit line will be consistent from one end of the line to the other. That is, not all close to the line at one end, and far off it at the other end. The last assumption is outliers.

Outliers are data points in the population that do not seem to fit the general trend of the other data points. Outliers can have a disproportionate influence on Pearson's correlation coefficient, *r*. There is a difference between extraneous outliers and a data set that has a natural skew. Dealing with a skew or extraneous data is always a judgment call (Warner, 2013). Some researchers prefer to remove high or low scores to avoid disproportionate impact on correlation results, while other researchers employ techniques such as data transformation of higher or lower scores, or separate analyses of the normal and non-normal portions of the distribution, or taking larger samples (Cohen, Cohen, West, & Aiken, 2003; Warner 2013). It may be unnecessary to be concerned about this skew. "The central limit theorem tells us that in big samples, the sampling distribution will be normal regardless… early research did indeed show that with samples of 40 the normality of the sampling distribution was … normal" (Field, 2009, p.156). In any case, outliers should be eliminated, transformed, or otherwise accounted for before sampling or analysis is conducted.

Sampling Method

After eliminating cases for missing or invalid data, eliminating or transforming outliers, and conducting assumption testing, the researcher proceeded with sampling. The sampling frame was accessible and convenient to the researcher and the sampling method was random with all valid records having an equal probability of selection. First, all of the data records were assigned an 8-digit random number from 0 to 1 using the un-seeded uniform random number generator in SPSS. For Research Question 1, the data was sorted in ascending order according to the random numbers and the first 100 cases were selected. The correlation analysis was conducted for RQ 1 and the results recorded. For RQ 2, the data was filtered to include only males in the cohort and then the first 100 cases were selected. The correlation analysis was conducted for RQ 2 and the results recorded. For RQ 3, the data was filtered to include only females and the first 100 cases were selected. The correlation analysis was conducted for RQ 3 and the results recorded. The hypothesis tests for these three samples were conducted to answer Research Questions 1, 2, and 3.

Research Questions

After the data was cleaned or adjusted as necessary and assumption testing was conducted, then sampling was conducted and the research questions were addressed.

Research Question 1

Research Question One was: Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average? The corresponding null hypothesis tested was: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average.

To test the null hypothesis for Research Question 1, a standard bivariate correlation was performed with IPI score as the predictor variable and FSGPA as the criterion variable. The significance level was set at $\alpha = .05$ with two tails. Correlational designs are often used when determining the predictive validity of HSGPA and SAT scores for college grade point averages (Bridgeman, Pollack, & Burton, 2008; Higdem, et al., 2016; Noble & Sawyer, 2004; Warne et al., 2014). This is an appropriate analysis model for examining the predictive relationship between two variables (Field, 2009; Gall et al., 2007).

Research Question 2

Research Question Two was: Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males? The corresponding null hypothesis tested was: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males. An analysis sequence the same as for RQ 1 was conducted, but was restricted to males in the cohort.

Research Question 3

Research Question Three was: Is there a significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females? The corresponding null hypothesis tested was: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females. An analysis sequence similar to RQ 1 was conducted, but restricted to the females in the cohort.

Summary

This correlational study was based on the creation of an index that measured performance relative to ability and testing how well this index might predict first-semester college grade point averages. The setting for this study was a mid-sized regional, liberal-arts college in the Northeast. The primary element of the instrumentation was the construction of the index for each student. The preliminary steps of data analysis included range restrictions, handling missing and outlier data points, taking precautions for a small skew in the predictor variable, and recoding some elements. The primary model for analysis was bivariate correlation. Testing of the research questions revealed differences in the data set due to sex and possibly due to various ranges of the data. Because of this, additional tests were conducted including examining the data separately by sex for various ranges of the predictor variable. This was done using bivariate correlation and simple linear models. This also included performing a stepwise multiple regression to estimate the predictive value of the IPI, SATC, and HSGPA for the criterion variable of FSGPA. Subgroups were developed according to sex, high vs low IPI, and top half vs bottom half SATC percentiles.

The primary tools for the analysis was IBM SPSS, Version 23. Some other tasks, including subgroup analysis of various specific statistics were completed with Microsoft Excel 2010. The statistics reported included Pearson Correlation, significance, and descriptive statistics of mean (*M*), standard deviation (*SD*), number (*N)*, degrees of freedom (*df*), *r*, R Square, Adjusted R Square, F-Change (*F*), unstandardized coefficient (*B*), and standardized coefficient (beta).

CHAPTER FOUR: FINDINGS

Introduction

The purpose of this nonexperimental, correlational study was to predict success in college by creating and examining a new metric from existing and commonly used measures. The new metric indexes each student's performance against that student's ability and then tests the correlation between that index and college grades. The student's ability is measured by cumulative SAT score (SATC). Performance is measured by high school cumulative grade point average (HSGPA). HSGPA is frequently used to predict college performance, but it is a bare measure that does not take into account the student's ability. Examining only the HSGPA affords little indication of what the student's native ability level might be. For example, a 3.0 GPA might be considered high performance, expected performance, or even low performance depending on the individual student's ability. The index, the Individual Performance Index (IPI), was created to factor in the student's ability before HSGPA was used as a measure of performance. This index was examined using bivariate correlation to observe the relationship between it and the criterion variable, which is first-semester college cumulative grade point average (FSGPA).

Data Preparation

In preparing the data for analysis, the data set was inspected and organized. The population initially consisted of 544 cases. After elimination of cases for missing or incomplete data 356 cases remained. For some cases, there were missing or incomplete SATC scores, but there was an ACT composite score. The ACT composite score along with the College Board Concordance Tables could have been used to impute the SATC score (Warner, 2013). However, this would have added only a few records (n = 12), so this decision was not taken. The demographics of the remaining cases were examined to determine if the population demographics changed significantly because of the eliminations. For males, 87 cases were removed, which brought the male percentage of the population from 43.6% (n = 237) to 42.1% (n = 150). For females, 101 cases were removed, which brought the female percentage of the population from 56.4% (n = 307) to 57.9% (n = 206). For whites, 157 cases were removed. This brought the percentage of white students from 88.6% (n = 482) to 91.0% (n = 324). For non-whites, 30 cases were removed for missing or incomplete data. This brought the percentage of non-white students from 11.4% (n = 62) to 9.0% (n = 32).

After eliminating cases for missing or incomplete data, the next step was recoding the HSGPA to a 4.0-5.0 scale where necessary. After this, the IPI was calculated for each student. To do this, all SATC and HSGPA scores were converted to z-scores and then to percentiles. The IPI was determined for each student by taking the quotient of HSPGA percentile and SATC percentile.

After these figures were computed, the data was examined for extraneous outliers or other anomalies. Eleven cases were eliminated when assumption testing was performed, as described below. This brought the sampling population to 345 cases. Random samples of 100 from the population were taken to test each of the three hypotheses.

Descriptive Statistics

The four variables involved in this study were the IPI, FSGPA, HSGPA, and SATC. Descriptive statistics for these variables were observed and the results briefly presented here. These are descriptives for the entire sampling population or a specific subpopulation to add clarity or information for some aspect of the distribution. The descriptives for each hypothesis test sample are included in the section with that hypothesis test.

Individual Performance Index

The Individual Performance Index (IPI) is constructed as the quotient of HSGPA percentile over SATC percentile. An IPI score of 1.0 indicates the HSGPA percentile matches the SATC percentile; i.e. that performance matches ability. An IPI score of, for example, 1.28 indicates performance that is 128% of ability. After eliminating outliers, the statistics for the IPI were $M = 2.09$ and $SD = 4.823$. There was a positive skew observed in the distribution of the IPI. In order to moderate the impact of the positive skew in the IPI distribution, all IPI scores were transformed (capped) at mean + 3SD or 16.5. After the transforming of the IPI, the descriptive statistics are $M = 1.79$, $SD = 2.905$, $N = 345$. The histogram distribution for the IPI after this transformation of the skew is in Figure 1. The Shapiro-Wilk statistic is $W = .463$ ($df = 345$, $p < .001$).

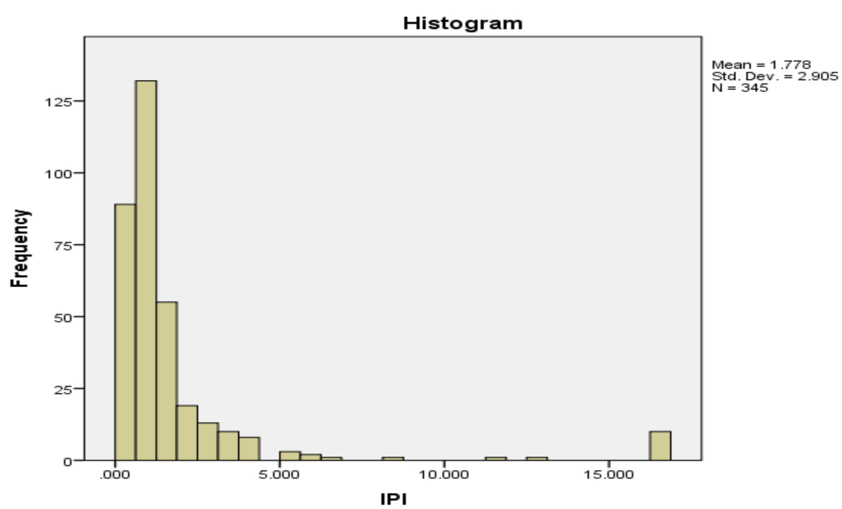

Figure 1. Histogram of IPI for Incoming Freshman Cohort (outliers capped at mean + 3SD)

It is clear from visual inspection this distribution has a positive skew. The cases in the skew represent students with very low SATC scores and much higher relative HSGPA. Figure 2 shows the Q-Q Plot for the IPI.

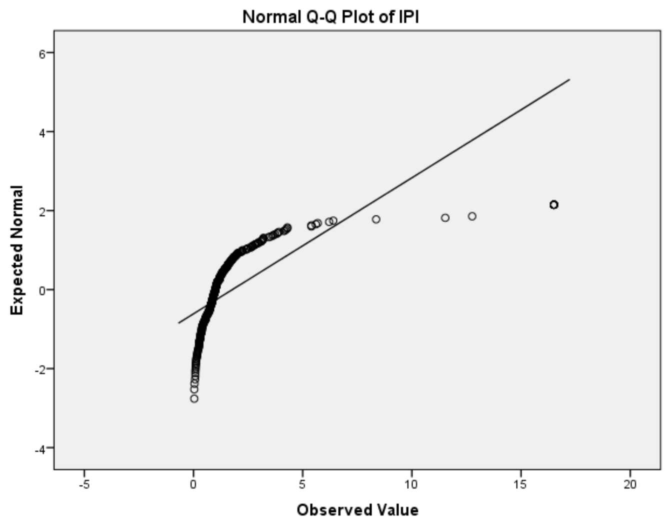

Figure 2. Q-Q Plot of IPI for Incoming Freshman Cohort (outliers capped at mean + 3SD)

Figure 3, below, shows a zoomed-in view of the IPI score distribution from 0.0 to 3.0. The data set is near-normal when the skew is eliminated. This portion of the sampling population represents 88.7% (*N* = 306) of the cohort.

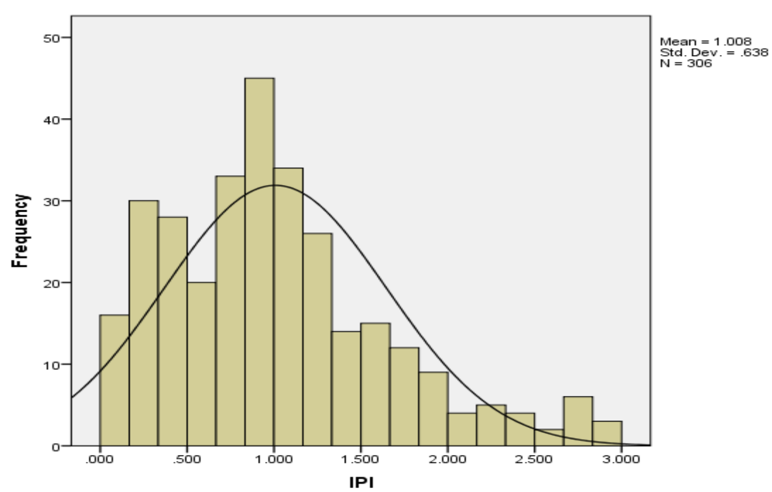

Figure 3. Histogram of IPI for Incoming Freshman Cohort with Range Restricted 0.0 to 3.0

Part of this study was to make separate comparisons of the relationship between IPI and FSGPA for each sex. Figure 4 shows the distribution of random samples of approximately 100 IPI scores for females and males with IPI range limited to 0.0 to 4.0. The distribution for females on the left is an approximate normal distribution with heavy concentration at 1.0 (*M* = 1.28, *SD* =

.771). The distribution for males on the right is non-normal with a higher concentration of scores below 1.0 (M = .891, SD = .826). The mean IPI scores between the sexes are significantly different (p < .01, two-tailed) and the Cohen's d effect size is 0.487. Cohen's d was calculated using the difference of the means divided by the pooled variance (Field, 2009; Warner, 2013).

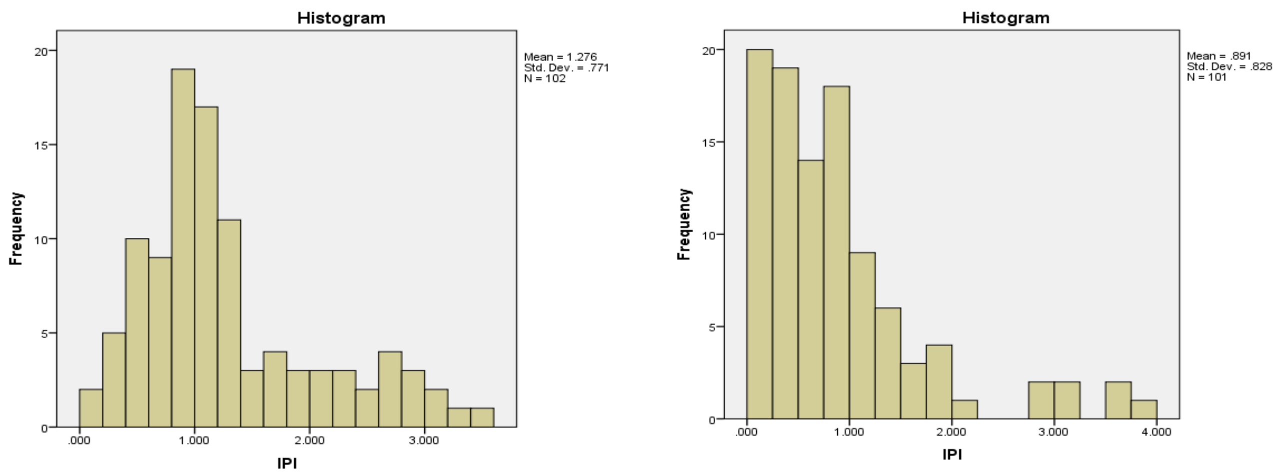

Figure 4. IPI Distributions for Females (left) and Males (right) in IPI Range from 0.0 to 4.0

First-semester Grade Point Average

The grading system at Northeast College is on a 4.0 scale. There is no weighting scheme. The histogram distribution for the first-semester college grade point average (FSGPA) for the entire incoming freshman cohort is in Figure 5. The Q-Q plot is in Figure 6.

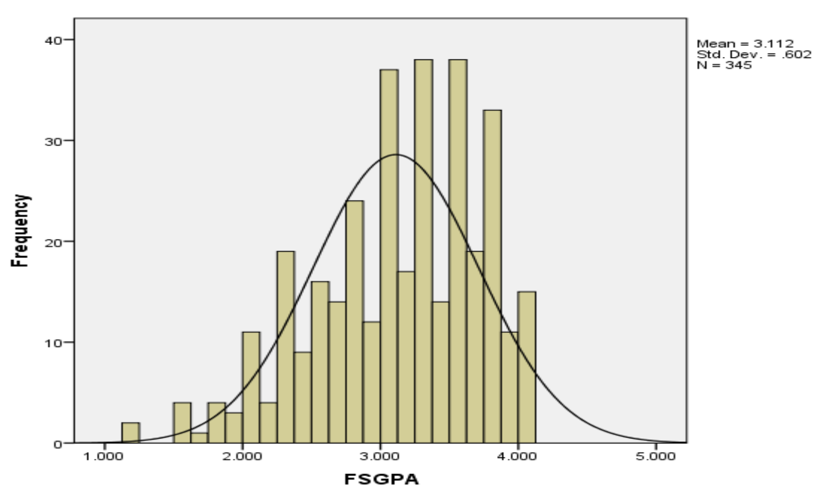

Figure 5. Histogram of FSGPA for Incoming Freshman Cohort

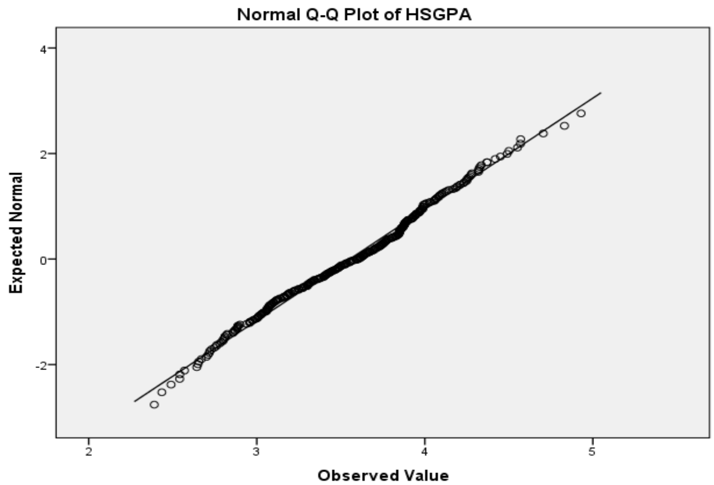

Figure 6. Normal Q-Q Plot of FSGPA for Incoming Freshman Cohort

The alternating longer bars of the histogram in Figure 4 are an artifact of the 11 discrete grades that can be earned in each of 4 classes in the first fall semester and their corresponding numeric equivalents. For example, a student earns a 3.7 for an A- or a 3.3 for a B+. It is not possible to earn any numeric value between 3.7 and 3.3 for a particular grade. This is relevant because it produces abnormally higher differences between the expected values (the superimposed normal curve) and the data (bars). This artificially inflates the S-W test results (because of the way S-W statistic is calculated) and renders it less useful ($W = .959$, $df = 345$, $p < .001$). Bivariate normality is already difficult to evaluate (Warner, 2013) and it was not easily evaluated here because of this and the skews found in the FSGPA and IPI distributions. The Q-Q plot indicates a normal distribution. Precautions were taken to reduce undue influences here, including using large samples sizes and transform (cap) some of the data in the positive skew for the IPI.

Part of the reason for the negative skew in FSGPA is the preponderance of high grades. Tucker and Court (2010) report that the "problem of grade inflation has been a topic of concern for over a century and there are no quick fixes or simple methods of reversing this trend" (p. 45).

This is an international concern and seen in the United Kingdom, Vietnam, and Hong Kong (Tucker & Court, 2010). The pressures for this inflation come from various sources including instructor leniency, faculty evaluations, loan or scholarship requirements, and rising expectations from employers (Butcher, McEwan, & Weerapana, 2014; Kostal, Kuncel, & Sackett, 2016; Tucker & Court, 2010). At Louisiana State University's School of Social Work in the Spring 2008 semester, over 79% of the grades awarded were As (Tucker & Court, 2010).

The distributions of FSGPA for females and males from the same samples as depicted in Figure 2 are shown in Figure 7.

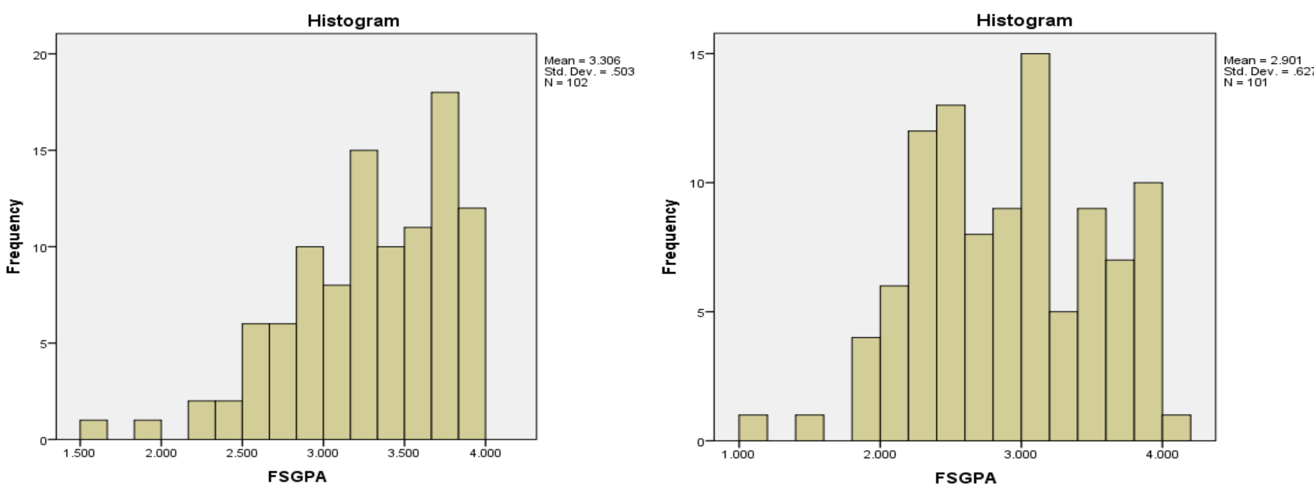

Figure 7. FSGPA Distributions for Females (left) and Males (right) in IPI Range from 0.0 to 4.0

The mean FSGPA for the females in this group is 3.31 with standard deviation of 0.503. The mean FSGPA for males is 2.90 with standard deviation of 0.627. These means are significantly different ($p < .01$, two-tailed) and Cohen's d is 0.721. This is a fairly large effect size (Cohen, 1988).

SATC and HSGPA

SATC and HSGPA are important to this study because they are the base measures from which the IPI is constructed. For the entire sampling population of 345, the mean for SATC for females and males are 1731 and 1728, respectively. The standard deviations are 179.4 and 163.3, respectively. These are not significantly different ($p > .05$, two-tailed). The similarity of SATC scores for females and males at Northeast College mirrors national trends (SAT, 2013b).

The means for HSGPA for females and males are 3.69 and 3.37, respectively. The standard deviations are 0.423 and 0.481, respectively. This difference is significant ($p < .01$, two-tailed) and Cohen's *d* effect size is 0.707. This indicates a fairly large difference between these means (Cohen, 1988). The average HSGPA is significantly higher for females and this corresponds to reports of national results (Cornwell, Mustard, & Van Parys, 2013).

Assumption Testing

Prior to conducting the bivariate correlation analyses, assumption testing was conducted. The assumption of normality for each variable was evaluated using histograms, Q-Q plots, and the Shapiro-Wilk test. These were presented above with the descriptive statistics overview. "With large enough sample sizes (> 30 or 40), the violation of the normality assumption should not cause major problems" (Ghasemi & Zahediasl, 2012, p. 486). Field (2009) also indicates that with sample sizes above 40, the normality of the sampling distribution will be assured, based on the central limit theorem. It is unnecessary to be concerned about skewness discussed below.

Scatterplots for the two study variables, IPI and FSGPA, were examined for linearity and homoscedasticity. These results are at Appendix C. Two views of each scatterplot were presented for each hypothesis test sample to overcome some limitations of scale and the mechanics of SPSS charting. The first view is the entire sample; the second view is an expanded view of the of the sampled test cases which makes a potential best it line easier to project. There is a slight problem with homoscedasticity for the RQ 1 sample, with slightly smaller variance present at the higher IPI end of the graph. The variance appears fairly consistent for the scatter plots for RQ 2 and RQ 3, although there are more data points at the smaller IPI end of the plots. There is not a highly correlated linearity for any of the plots, i.e. where the points are very close to a particular best fit line. However, there also does not appear to be a non-linear shape to the distribution of scatterplots either. It appears from these plots that the assumptions of linearity and homoscedasticity are not unreasonably violated.

Extreme Outliers Excluded

After calculating the index and before running the analyses, a small number of outliers were identified. Extreme outliers were identified using histograms, boxplots, and z-scores (distance from the mean in standard deviations). To protect the correlation assumption of no extreme outliers, cutoffs of more than 3.0 standard deviations from the mean for either SATC, HSGPA, or FSGPA were established. A cutoff of below 1st percentile for SATC or HSGPA was also used (Field, 2009; Warner, 2013). A total of 11 cases were eliminated as outliers. This was 3.1% of the population and brought the final sampling population to 345.

Outliers versus Skewed Data Elements

After extreme outliers were excluded, the IPI was calculated for the remaining population cohort (n = 345). A positive skew was apparent in the distribution of the IPI. To limit any undue influence of this positive skew, the data points in the skew were transformed by capping them at mean + 3SD. The decision was made to transform, rather than eliminate, the positive skew. On inspection of the 12 highest IPI scores, all of these students had a HSGPA of at least 3.1 and all but 3 of them were above 3.5. However, they also had an SATC score in the single digit percentiles within their freshman cohort. This combination is what raised their IPI index. The data points in the skew of the IPI distribution represented reasonable cases which are germane to the purpose of this study. Because of this, instead of eliminating them as extreme outliers, the decision was taken to mitigate any negative impact by taking two precautions. The first precaution was to use transformation to cap the higher IPI indices at the mean plus 3.0 SD. Because Pearson's r is not robust for correlational studies with a small sample size, the second precaution was to use a sample size of 100 (Ghasemi & Zahediasl, 2012; Warner, 2013). These precautions help to limit negative influence on the statistical models, but still account for these relevant data points in the analysis (Field, 2009). Because the assumptions did not appear to be violated and because precautions were taken, the decision was made to proceed with analysis and conduct the bivariate correlations as planned.

Hypotheses Test Results

Null Hypothesis 1

Null hypothesis 1 stated: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average. It was tested using a bivariate correlation analysis. The incoming freshman cohort ($N = 345$) was randomly sampled for 100 cases. This sample included 62 (62%) females, 38 (38%) males, 6 (6%) Hispanic, and 11 (11%) non-White students. The bivariate correlation was performed with the two variables of FSGPA and IPI on the sample of 100. The results of the bivariate correlation between IPI ($M = 1.61$, $SD = 2.46$, $N = 100$) and FSGPA ($M = 3.09$, $SD = .600$) were $r(98) = -.107$, $p = .291$. There was no significant correlation between the between the Individual Performance Index and first-semester college grade point average because the p-value was greater than .05. Null hypothesis one was not rejected.

Null Hypothesis 2

Null hypothesis 2 stated: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for males. It was tested using a bivariate correlation analysis. The incoming freshman cohort was randomly sampled for 100 males. The results of the bivariate correlation between IPI (M = 1.37, SD = 2.57, N = 100) and FSGPA (M = 2.94, SD = .627) were $r(98) = .042$, $p = .677$. These results indicated there is not a significant correlation between the Individual Performance Index and first-semester college grade point average for males ($p > .05$). Null hypothesis two is not rejected.

Null Hypothesis 3

Null hypothesis 3 stated: There is no significant predictive relationship between the Individual Performance Index and first-semester college grade point average for females. The incoming freshman cohort was randomly sampled for 100 females. The bivariate correlation was performed with the two variables of IPI and FSGPA. The results of the bivariate correlation between IPI (M = 1.95, SD = 2.87, N = 100) and FSGPA (M = 3.25, SD = .571) were $r(98) = –.369$, $p < .001$. These results indicated that there is a significant relationship between the Individual Performance Index and first-semester college grade point average for females ($p < .001$). Null hypothesis three is rejected. The Pearson coefficient of –.369 is a medium negative effect size. The negative element of this means that IPI and FSGPA are inversely related; as an individual's IPI score increases, the FSGPA decreases.

Additional Analysis

Linear Regression Model

HSGPA and SATC scores are the most common predictors used for college grades (Gaertner & McClarty, 2015). A standard multiple linear regression was performed to show the predictive relationship of a model predicting FSGPA from HSGPA and SATC. To mirror the hypotheses and previous results in this study, the regression was run for three groups: mixed sex, females, and males. Table 4 shows these results the same samples that taken for the hypotheses tests. Because the scale of SATC scores (eg. 1700) is considerably larger than HSGPA (eg. 3.25), standardized coefficients are reported.

Table 4

Multiple Linear Regression Predicting FSGPA from SATC and HSGPA

RQ	N	Model Predictors	R Sq.	Adj R Sq.	Coeff	Std Coeff	Sig	VIF	Durbin-Watson
1	100	SATC	.401	.389	.001	.315	.000		
		HSGPA			.592	.447	.000	1.155	2.019
2	100	SATC	.354	.341	.001	.158	.085		
		HSGPA			.661	.509	.000	1.235	1.802
3	100	SATC	.441	.429	.002	.520	.000		
		HSGPA			.350	.250	.003	1.207	2.049

CHAPTER FIVE: DISCUSSION, CONCLUSIONS, AND RECOMMNEDATIONS

Introduction

The purpose of this non-experimental, correlational study was to create an index that reports students' academic performance relative to their ability and then to assess the correlation of the index with first-semester college grade point average (FSGPA). Additional hypotheses tests were conducted to examine whether this index had different correlation values when the data was disaggregated by sex. The basis for examining males and females separately in this study are the documented discrepancies between the sexes in noncognitive characteristics and behaviors (Koul, Roy, & Lerdpornkulrat, 2012; Weis, Heikamp, & Trommsdorff, 2013).

The researcher sought to understand if making grades relative to ability might be a viable predictor of FSGPA. The index was created from the two most common predictors of success in college: high school grades and standardized test scores (Bai, Chi, & Qian, 2014; Bettinger, Evans, & Pope, 2013; Crede et al., 2010; Ledsema & Obukova, 2015). The name used for this index was the Individual Performance Index (IPI). In this chapter, the researcher discusses the findings of the study against the backdrop of the literature and theory. The chapter also includes research implications, practical implications, limitations of the study, suggestions for further research, and conclusions. It is important to keep in mind the population in this study consisted only of college-bound individuals, with attendant motivations and ambitions. The results of this study likely do not extend to the general body of high school students and should be applied with caution. The primary intended audiences for this study were admissions and student affairs personal.

Discussion

The researcher used a convenience sample of a freshman cohort from a 4-year private liberal arts college. The college is categorized as having "moderately difficult" admission standards according to Peterson's Publishing; the Carnegie enrollment profile for this college is classified as "Very High Undergraduate." The anonymous, archival data was provided to the researcher by the college. Cases with missing or incomplete data were removed. Precautions were taken to reduce the impact of any extraneous data.

The construction of the IPI was completed using the archival data and is based on two theoretical frameworks. The first is indexing (Fischer, 1923). The function of indexing in this study is to render measures more meaningful by putting them in the context of a relevant benchmark. Here, grades (i.e., HSGPA) are examined against the relevant benchmark of ability (i.e., SATC). The second theoretical framework is the emerging research focusing on the importance of noncognitive factors in educational success (Gaertner & McClarty, 2015; Lipnevich & Roberts, 2012; Sohn, 2010). Success in college can be attributed to both cognitive and noncognitive factors, and noncognitive factors influence grades more than tests of ability (Duckworth, Quinn, & Tsukayama, 2012). Economists and psychologists are starting to realize that noncognitive factors highly influence success, even more than IQ (Tough, 2012).

There is not a lot of quantitative research directed toward noncognitive factors as predictors of academic success because noncognitive factors can be difficult to isolate and quantify (Van Ark, 2012). This is beginning to change (Lipnevich & Roberts, 2012; Thomas, Kuncel, & Credé, 2007). A difficulty for examining noncognitive factors is the inherent amorphous and intertwined nature of various noncognitive phenomena (Borghans, Duckworth, Heckman, & Weel, 2008; Farrington et al., 2012; Willingham, 2013). They are hard to isolate and hard to quantify. However, the "power of using indices as management tools clearly resides in their ability to capture the information contained in a large number of variables in one number" (U.S. Department of Energy, 1995, p. 1–59). Another benefit of indices is to quantify phenomena that are essentially qualitative (Jaschik, 2008; Gillum, 2012) . This study created a metric that indexed a measure with a high noncognitive component (HSGPA) with a measure that does not contain a high noncognitive component (SATC). This isolation and quantification of noncognitive influence in performance is the basis of the IPI. This quantification of noncognitive influence is based on an amalgamation of noncognitive phenomena, rather than isolation of individual elements, such as grit, self-regulation, and attitude. The index provided quantitative evidence of noncognitive differences between the sexes and showed how that evidence can impact predictors of academic success.

There were three research hypotheses in this study. The first examined the predictive relationship between the IPI and FSGPA for the incoming cohort. The second and third hypotheses examined the predictive relationship between the IPI and FSGPA separately for males and females.

The first hypothesis test showed no significant correlation between the IPI and FSGPA for the cohort (both sexes were included). The second hypothesis test also showed no significant correlation between the IPI and FSGPA for the sample of 100 males. These findings do not support or refute reports from the literature, except perhaps that for males, positive noncognitive behavior is less evident (Cleveland, 2011; Hickey, 2008; Neu and Weinfeld, 2007) and therefore does not impact high schools grades in the same positive manner as females (Connell & Messerschmidt, 2005; Weis et al., 2013). This is also suggested by the regression results where the model for males explains 35.4% of the variation in college grades, where the model for females explains 44.1%. The impact of these variables for males is less predictable.

The third hypothesis test had a different result. The bivariate correlation on the sample of 100 females did show a significant relationship between IPI and FSGPA. The relationship was significant, but it was a negative correlation. This means, in general, the more females exceed their expected performance in high school (i.e., as IPI goes up), the lower FSGPA they attain. If their grades remained high in college, this negative correlation would not appear. This may happen because reasons for good grades in high school may not be present to the same degree in college, including perhaps the impact of noncognitive factors. These results are consistent with reports from the literature and the developing noncognitive research that grades are more influenced by noncognitive factors (Duckworth, Quinn, & Tsukayama, 2012) and that females exhibit higher levels of positive noncognitive behavior in academic settings (Koul, Roy, & Lerdpornkulrat, 2012; Weis, Heikamp, & Trommsdorff, 2013).

This study has shown that examining subgroups of cases can provide useful, actionable information. These results provide a cautionary tale for admissions and counseling personnel that students who overperform in high school, may not continue that trend in college. The significant negative correlation of the IPI to FSGPA for females indicates that performing above relative ability in high school may not be replicated in college. This is consistent with the regression model where females' high school grades had a coefficient of 0.350, where for males the high school grades coefficient was 0.661. High school grades were a less important predictor of college performance for females.

The distributions of HSGPA and SATC in this study were consistent with national reports that high school grades were higher for females and that SATC scores were similar for both sexes (Cornwell, Mustard, & Van Parys, 2013; SAT, 2013b). The result that females had higher IPI scores is consistent with the literature in that females exhibit more positive noncognitive behavior and that noncognitive behavior has a greater impact on grades (Duckworth, Quinn, & Tsukayama, 2012; Gaertner & McClarty, 2015; Koul, Roy, & Lerdpornkulrat, 2012; Weis, Heikamp, & Trommsdorff, 2013). When the data was not disaggregated, there were no significant findings, which indicates that examining only aggregated data could mask important information. When the data was disaggregated by sex a significant finding emerged.

Research Implications

The findings of this study were consistent with national reports, past research, and the literature. The findings also suggest ways to continue or improve research or statistical analyses for college personnel in some areas.

There is not a lot of quantitative research on noncognitive factors as predictors of academic success because noncognitive factors can be difficult to isolate and quantify (Lipnevich & Roberts, 2012; Thomas, Kuncel, & Credé, 2007; Van Ark, 2012). Noncognitive phenomena are intertwined and do not have clear boundary conditions where one of them starts or stops (Borghans et al., 2008; Farrington et al., 2012; Willingham, 2013). However, noncognitive factors are beginning to play a central role in forming education policy around the world (Lipnevich and Roberts, 2012). Combining these educational impetuses with indices that capture amalgamations of noncognitive factors can create powerful tools for analysis (U.S. Department of Energy, 1995). This study developed a metric that revealed one significant finding. Since noncognitive measures are likely to defy clear isolation and quantification, this kind of juxtaposing of common measures that include different degrees of noncognitive and cognitive influence could help the effort to improve quantitative analysis of noncognitive phenomena. The result here is quantitative tool that can predict college success in at least one common scenario. It is also seemingly reasonable to develop tools that are not so complicated as to defy any easy interpretation at the practical, school-based level. The IPI tool developed here enjoys the benefits of being easy to calculate, readily available to colleges, and statistically significant in the case of females.

Practical Implications

The usefulness of the IPI and the indexing process developed in this study may have practical implications. Because the results here indicated significance for one subgroup and not for others, admissions and student affairs personnel might reexamine their current modeling methods for improvement when subgroups of their populations are processed separately. It is possible that current models are better than indicated when used on particular subgroups, or less useful for other subgroups.

The IPI allows students with lower ability to compete with students of higher ability on a more level playing field. A student who improves their IPI from 95% to 110% (.95 to 1.1) can do this without changing their standardized test scores, which is difficult to do. This allows for quantifying, identifying, and rewarding improved noncognitive behavior. This can give context and encouragement to individuals or groups of students no matter where their ability levels lie. The IPI can also identify students with high ability, typically males in this study, who have only decent grades when they really should have superior grades.

The IPI may be a more equitable way to compare students for awards or athletic eligibility. One of the complaints of the NCAA minimum GPA eligibility requirements is that it is an unfair standard for students with lower native ability. Because the IPI accounts for differences in ability, it may be a more equitable criteria for awarding scholarships or making NCAA eligibility decisions, rather than grades alone.

For some students a high IPI, which indicates high performance in high school, is not followed by high grades in college. This was a significant general finding for females in this study, but it can also be identified on an individual basis. Student support services can proactively identify students who may be feeling disheartened because their overperformance in high school is not continuing in college and provide proactive counseling. Since SAT exams are often taken in junior year of high school, the high school counselors could construct and use the IPI for individual students as part of the counseling or guidance process.

Limitations

A primary limitation of the IPI is that its construction is based partly on high school grades and HSGPA. Individual grades and grade averages are often rough instruments at best. The vagaries of grading and grading systems used in high schools include differences in curriculum, courses taken, high school cultures, teachers, GPA computation systems, and various obvious and subtle biases. According to the admissions director of Northeast College, this is simply a fact of life and is not likely to go away or improve much. Nonetheless, HSGPA remains a statistic they obtain whenever possible and it impacts their process.

While IPI scores, especially if tracked over time, might prove very useful to a specific college, it is not clear that IPI values would travel well between colleges. The demographics of a particular college and its admissions profile, and the college itself, might prove a serious disclaimer to any interpretations of their data. SATC scores are standardized, but external validity must be disclaimed and different schools should examine their own procedures to determine the most applicable data to use in developing their models.

There was no tracking of course selection either in high school or college in this study. Some students take easier tracks in high school and easier or harder tracks in college, for example STEM curriculums, and these factors are not recognized in this study. If the data on specific tracks is available, it might provide a separate, useful construction of the IPI, but it was not part of this study. It is another way to disaggregate the data which might prove useful.

Lastly, the IPI exhibited a positive skew. This was addressed by precautions, but repeating the analyses and restricting the skew to mean + 1.0 SD might provide more confidence that the skew did not overly influence the results. Holdout samples might also be tested to see if they confirm the initial results. This is less likely given the sample sizes of 100, nearly 1/3rd of the sampling population, but holdout samples might improve confidence that the random sampling itself did not create unrepresentative results.

Recommendations for Future Research

The emerging research and importance of noncognitive behavior in academic performance warrants that any effort to quantify noncognitive behavior be pursued. Mechanisms for measuring various noncognitive phenomena are already being developed, for example, the Grit Scale developed by Duckworth et al. (2007). Better mechanisms for isolating and measuring specific noncognitive phenomena might reveal the particular noncognitive elements, eg. grit or self-regulation or self-efficacy, that have the greater impact on academic success. In the meantime, the IPI is easily calculated from readily available and ubiquitous measures. It can be studied in myriad environments, formulations, or subgroups. Future research might include more robust sampling populations, modifications of which particular grades or tests are used to calculate the IPI, disaggregation for multiple subgroups, or gauging the impact if it replaced or was added to HSGPA to determine scholarships or athletic eligibility.

This study has confirmed that disaggregation of data can be informative. The IPI can also be modified to examine particular sub-measures of performance or ability. For example, a school district could examine the index of 8th grade math grades to a standardized state-mandated mathematics test for particular groups of students. They could use the index to reveal if a significant group of students have poor grades although they have high test scores in the same subject area. Incongruent results might trigger closer examination of curriculum, grading bias, or test validity.

The mechanism underlying this study, the IPI, which relativizes performance to ability, could be reconstituted and/or replicated in a multitude of scenarios. Some of the ways to do this could address limitations of this study. Restricting the categories of grades included both from high school and college could reduce the problems associated with crossing disciplines and lower, or higher, grades being a function of the curriculum, not the performance. For example, only STEM curriculum grades might be used to create and track a STEM-IPI. The construction of the STEM-IPI might also include only the MATH SAT score as the benchmark, instead of the composite SATC. This might be informative or useful given the emphasis toward overcoming sex enrollment imbalances in STEM career fields. A complementary NON-STEM-IPI could be also calculated with curriculum that does not fit into STEM. Analyses across college consortiums or between colleges might improve understanding of how and where the IPI, or a variation of it, or a similar instrument, might be useful. Various creative disaggregations of the data might also reveal significant differences between subgroups, or reveal that there are statistically significant subgroups where none were previously identified. For example, as advised by Cohen et al. (2003), disaggregation based on various ranges of variables might produce significant and actionable information.

If the IPI were computed statewide, it might provide a state community college system with a mechanism to track and evaluate their own efforts at improving noncognitive behavior among their students or provide comparisons with other state community colleges. Of course, this is not restricted to 2-year schools. Because noncognitive behavior has been shown to heavily impact academic success, tracking noncognitive improvements could provide feedback and incentives to colleges and students.

For regional colleges, who enroll many dozens of students from the same high schools every year, tracking the IPI scores and college success for their cohorts might alert the admissions and student support personnel of issues with HSGPA within certain high schools. College personnel are already aware of differences in high schools and grades failing to be replicated in college. The IPI offers a mechanism to view that quantitatively. The IPI could be used as a longitudinal measure for the college itself. Once students are enrolled in a particular college, some environmental confounders from their varied pre-college backgrounds are controlled for. The college could track whether the IPI score of individuals, cohorts, or various sub-populations is rising, declining, or unchanged; and whether the IPI subsequently serves as a predictor of future success, retention, or completion. The current results only indicate significant findings for females in this particular cohort. Further research might produce useful findings for other subgroups.

The IPI could be constructed from other noncognitive measures instead of HSGPA. For example, a measure of grit or self-regulation or other noncognitive measure could be used in the numerator to see if that new configuration provided useful information. Once a student is enrolled in a college, the numerator of the IPI might be replaced with college grades. The index might also be studied for its ability to predict degree completion, time to completion, or retention.

Conclusions

The results of this quantitative study are consistent with the research and literature that noncognitive behavior influences academic performance and that females exhibit better noncognitive behaviors overall. The results also indicate that performing well above ability in high school for females is not necessarily replicated in college. Implementing quantitative tools such as indexing can sometimes produce significant findings, even in the context of relativizing difficult-to-quantify concepts like noncognitive behavior. Disaggregation of data is shown to be a potentially useful method of analyses and capable of revealing actionable information that aggregated reporting obscures. Disaggregation in this study data revealed different results for females and males. Finally, this study showed that using the same modeling procedures or parameters to predict college success for all subgroups may not be a sound policy.

REFERENCES

ACT (n.d.a). Compare ACT & SAT scores. Retrieved from http://act.org/solutions/college-career-readiness/compare-act-sat/

ACT (n.d.b). Should I test again? Retrieved from http://www.actstudent.org/faq/more.html

ACT (2007). The ACT Technical Manual (2007). Retrieved from http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf

ACT (2013a). ACT Profile Report – National. Retrieved from http://www.act.org/newsroom/data/2013/pdf/profile/National2013.pdf

ACT (2013b). National Ranks for Test Scores and Composite Score. Retrieved from http://www.actstudent.org/scores/norms1.html

ACT (2013c). National Score Trends. Retrieved from http://www.act.org/newsroom/data/2013/trends.html

ACT (2013d). The Condition of College and Career Readiness – Asian Students. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/CCCR-2013-Asian.pdf

ACT (2015). The Condition of College and Career Readiness – National ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/Condition-of-College-and-Career-Readiness-Report-2015-United-States.pdf

Ainley, M., Hillman, K., & Hidi, S. (2002). Gender and interest processes in response to literary texts: Situational and individual interest. *Learning and Instruction, 12*, 411–428.

Bacon, D. R., & Bean, B. (2006, April). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education, 28*(1), 35-42.

Balsa, A. I., Giuliano, L. M., & French, M. T. (2011, February). The effects of alcohol use on academic achievement in high school. *Economics of Education Review, 30*(1), 1–15.

Bai, C., Chi, W., & Qian, X. (2014, September). Do college entrance examination scores predict undergraduate GPAs? A tale of two universities. *China Economic Review, 30*, 632–647.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1989, September). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184. doi:10.1037/0003-066X.44.9.1175

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization, 6*(3): 285–327. doi:10.1177/1476127008094339

Bell, E. C., Willson, M. C., Wilman, A. H., Dave, S., & Silverstone, P. H. (2006, April). Males and females differ in brain activation during cognitive tasks. *NeuroImage, 30*(2), 529–538. doi:10.1016/j.neuroimage.2005.09.049

Berry, C. M., & Sackett, P. R. (2009, July). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science, 20* (7), 822-30.

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2013, May). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy 2013, 5*(2), 26–52. doi:10.1257/pol.5.2.26

Bhanot, R. T., & Jovanovic, J. U. (2009). The links between parent behaviors and boys' and girls' science achievement beliefs. *Applied Developmental Science, 13*(1), 42–59.

Birmingham Post (1999, April 24). Making reading an adventure (News). Retrieved from https://www.highbeam.com/doc/1G1-60542733.html

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Borghans, L., Duckworth, A., Heckman, J. J., & Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources, 43*(4), 972–1059.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). Test scores and high school grades as predictors. In W. G. Bowne, M. M. Chingos, & M. S. McPherson (Eds.), Crossing the finish line: Completing college at America's public universities, pp. 112–133. Princeton, NJ: Princeton University Press.

Brannian, J. D., Schmidt, S. M., Kreger, D. O., & Hansen, K. A. (2001). Baseline non-fasting serum leptin concentration to body mass index ratio is predictive of IVF. *Human Reproduction, 16*(9), 1819-1826.

Bråten, I., Strømsø, H. (2008, April). Predicting achievement goals in two different academic contexts: A longitudinal study. *Scandinavian Journal of Educational Research*, 50(2), 127-148. doi:10.1080/00313830600575932

Bridgeman, B., Pollack, J., & Burton, N. (2008). Predicting grades in college courses: A comparison of multiple regression and percent succeeding approaches. *Journal of College Admission, 199*, 19-25.

Briggs, D. C. (2009). *Preparation for college admission exams.* Arlington, VA: National Association for College Admission Counseling.

Brown, S. D., Tramayne, S., Hoxha, D., Telander, K., Fan, X., & Lent, R. W. (2008). Social cognitive predictors of college students' academic performance and persistence: A meta-analytic path analysis. *Journal of Vocational Behavior, 72*(3), 298–308

Brubaker, R. (2016). The Dolezal affair: Race, gender, and the micropolitics of identity. *Ethnic and Racial Studies, 39*(3), 414-448. doi:10.1080/01419870.2015.1084430

Bull, R., Cleland, A. A., & Mitchell, T. (2013). Sex differences in the spatial representation of number. *Journal of Experimental Psychology: General, 142*(1), 181-192. doi:10.1037/a0028387

Burton, L. A., Henninger, D., & Hafetz, J. (2005). Gender differences in relations of mental rotation, verbal fluency, and SAT scores to finger length ratios as hormonal indexes. *Developmental Neuropsychology, 28*(1), 493–505.

Bussey, K., & Bandura, A. (1999, October). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676-713. doi:10.1037/0033-295X.106.4.676

Butcher, K., McEwan, P., & Weerapana, A. (2014). The effects of an anti-grade-inflation policy at Wellesley College. *The Journal of Economic Perspectives, 28*(3), 189-204. Retrieved from http://www.jstor.org/stable/23800582

Casillas, A., Robbins, S., Allen, J., Kuo, Y., Hanson, M. A., & Schmeiser, C. (2012, May). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology, 104*(2), 407-420. doi:10.1037/a0027180

Carter, J. D., & Narramore, B. (1979). *The Integration of Psychology and Theology.* Zondervan Corporation: Grand Rapids, MI.

Cattell, R. B. (1963) Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22. doi:10.1037/h0046743

Certo, S. T., Busenbark, J. R., Woo, H.-s., & Semadeni, M. (2016, February). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal.* doi:10.1002/smj.2475

Chaput de Saintogne, D. M., & Dunn, D. (2001). Gender and achievement in clinical medical students: a path analysis. *Medical Education, 35*(11), 1024–1033.

Cheng, W., & Ickes, W. (2009, December). Conscientiousness and self-motivation as mutually compensatory predictors of university-level GPA. *Personality and Individual Differences, 47*(8), 817–822.

Cleveland, K. P. (2011). *Teaching boys who struggle in school : Strategies that turn underachievers into successful learners*. Alexandria, VA: Association for Supervision & Curriculum Development.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*(6), 426-443.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen J., Cohen P., West S. G., & Aiken L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.

College Board (n.d.). Retaking the SAT. Retrieved from http://professionals.collegeboard.com/testing/sat-reasoning/scores/retake

College Board (2010). *The SAT program handbook 2009-10.* Retrieved from http://www.collegeboard.com/prod_downloads/sat/sat-program-handbook.pdf

College Board (2013). *Test characteristics of the SAT®: Reliability, difficulty levels, completion rates, January 2012–December 2012*. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/Test-Characteristics-of-SAT-2013.pdf

Conard, M. A. (2006, June). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*(3), 339–346.

Concordia University of Nebraska (n.d.). Concordia scholarships, awards and grants. http://www.cune.edu/admission/undergraduate/financial-aid/scholarships/

Conley, D. T. (2013, January 23). What's in a name? Rethinking the notion of 'noncognitive'. *Education Week, 32*(18), 20-21. Retrieved from http://www.edweek.org/ew/articles/2013/01/23/18conley.h32.html

Connell, R. W,. & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society, 19*(6), 829-859.

Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources, 48*(1), 236 - 264

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual.* Odessa, FL: Psychological Assessment Resources Inc.

Covalent (n.d.). *Developing Performance Indices.* Retrieved from http://support.covalentcpm.com/CustomerDocs/TT5_DevelopingPerformanceIndicies.pdf

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing g. *Intelligence, 36*, 719–729.

Cox, L. (2012, February 9). 5 experts answer: Can your IQ change? *Livescience.* Retrieved from http://www.livescience.com/36143-iq-change-time.html

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272–295.

Cramer, P. (2006). *Protecting the Self: Defense mechanisms in action*. New York, NY: Publisher: Guilford Press

Curie, L. K., Pisarik, C. T., Gintner, E. J., Glauser, A. S., Hayes, C., & Smit, J. C. (2012). Life-skills as a predictor of academic success - an exploratory study. *Psychological Reports 2012, 111*(1), 157-164.

DeAngelis, S. (2003). Noncognitive predictors of academic performance: Going beyond the traditional measures. *Journal of Allied Health, 32*(1), 52-57.

Daniels, L. M., Haynes, T. L, Stupnisky, R. H., Perry, R. P., Newall, N. E., & Pekrun, R. (2008). Individual differences in achievement goals: A longitudinal study of cognitive, emotional, and achievement outcomes. *Contemporary Educational Psychology, 33,* 584–608. doi:10.1016/j.cedpsych.2007.08.002

Daphna, J. (2011, September). Male or Female? Brains are Intersex. *Frontiers in Integrative Neuroscience, 5*(57), 1–5. doi:10.3389/fnint.2011.00057

Darnon, C., Muller, D., Schrager, S. M., Pannuzzo, N., & Butera, F. (2006, November). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology, 98*(4), 766-776. doi:10.1037/0022-0663.98.4.766

Dash, S. (2011). Metacognition in girls. *Social Science International, 27*(2), 211-218.

Davidson, H., Cave, K. R., & Sellner, D. (2000, April). Differences in visual attention and task interference between males and females reflect differences in brain laterality. *Neuropsychologia, 38*(4), 508–519. doi:10.1016/S0028-3932(99)00084-6

Dechter, G. (2007, January 11). SAT scores well in predicting college success: State panel finds test is accurate gauge of retention, graduation rates in Md. *The Baltimore Sun.* Retrieved from http://articles.baltimoresun.com/2007-01-11/news/0701110042_1_standardized-test-scores-college-success-variables

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012, June). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods, 44*(2), 587–605. doi:10.3758/s13428-011-0160-2.

Dijkstra, A. G. (2002). Revisiting UNDP's GDI and GEM: Towards an Alternative. *Social Indicators Research, 57*(3), 301-338.

Drexler, p. (2013, January 7). Women need more than confidence to succeed, they need ambition. *Forbes.* Retrieved from http://www.forbes.com/sites/peggydrexler/2013/07/15/women-need-more-than-confidence-to-succeed-they-need-ambition/

Driessen, G. & van Langen, A. (2013, June). Gender differences in primary and secondary education: Are girls really outperforming boys? *International Review of Education, 59*(1), 67-86.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939-944.

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*(1), 198-208. doi:10.1037/0022-0663.98.1.198

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087–1101.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment, 91*(2), 166–174.

Duckworth, A.; Quinn, P. D.; Tsukayama, E. (2012, May). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology, 104*(2), 439-451. doi:10.1037/a0026280

Fannon, S. A., Vidaver, R. M., & Marts, S. A. (2002, June). Research update: Sex, genes and hormones. *Trends in Cognitive Sciences, 6*(6), 230–231. doi:10.1016/S1364-6613(02)01914-9.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N.O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago: University of Chicago Consortium on Chicago School Research.

Feng, S., & Xue, S. (2014, February). Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition. *Annals of the Institute of Statistical Mathematics, 66*(1), 121-140.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.

Figge, F., Hahn, T., & Schaltegger, S. (2001, November). *Stakeholder Valued Added*. United Nations Center for Stability Management. Retrieved from http://www.un.org/esa/sustdev/sdissues/technology/stakeholdervalue.pdf

"Financial ratios" (2011). In V. L. Burton, III (Ed.), *Encyclopedia of small business* (4th ed., Vol. 1, pp. 578-581). Detroit, MI: Gale.

Fischer, I., (1923). *The making of index-numbers: A study of their varieties, tests, and reliability.* Boston, MN: Houghton Mifflin Company.

Fitzclarenge, L. (1999, March). Educating boys in sport and physical education: Using narrative methods to develop pedagogies of responsibility. *Sport, Education and Society, 4*(1), 51-62.

Flynn, J. R. (1999, January). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54*(1), 5–20. doi:10.1037/0003-066X.54.1.5

Fortin, N. M., Oreopoulos, P., & Phipps, S. (2013). Leaving boys behind: Gender disparities in high academic achievement. *National Bureau of Economic Research (NBER) Working Paper Series*, w19331. Retrieved from http://economics.yale.edu/sites/default/files/fortin-121108.pdf

Frey, M. C., & Detterman, D. K. (2004, June). Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science, 15*(6), 373-388. doi:10.1111/j.0956-7976.2004.00687.x

Gaertner, M. N., & McClarty, K. L. (2015). Performance, perseverance, and the full picture of college readiness. *Educational Measurement: Issues and Practice, 34*, 20–33. doi:10.1111/emip.12066

Gagné, F., & St Père, F. (2002, January–February). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30*(1), 71–100. doi:2048/10.1016/S0160-2896(01)00068-X

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). New York: Allyn & Bacon.

Giedd, J. N., Raznahan, A., Mills, K. L., & Lenroot, R. K. (2012). Review: magnetic resonance imaging of male/female differences in human adolescent brain anatomy. *Biology of Sex Differences, 3*(19), 1-9. doi:10.1186/2042-6410-3-19

Geist, E. A., & King, M. (2005). Different, not better: Gender differences in mathematics learning and achievement. *Journal of Instructional Psychology, 35*(1), 43-52.

Ghasemi A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology & Metabolism, 10*(2), 486-9. doi:10.5812/ijem.3505

Gilligan, C. (1982). *In a different voice.* Cambridge, MA: Harvard University Press.

Gillum, B. (2012, November14). *Return on employees.* Enterprise Worldwide. Retrieved from http://www.enterpriseworldwide.org/blog/?p=622

Gurian, M. & Stevens, K. (2005). *The minds of males: Saving our sons from falling behind in school and life*. San Francisco, CA: Jossey-Bass.

Hanson, C. C., & Zambo, D. (2010). Boys in the club: Exploring positive male archetypes with preschool males. *Literacy Research and Instruction, 49*, 40–55.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J., (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*(3), 562–575. doi:10.1037//0022-0663.94.3.562

Hasselhorn, H.-M., Müller, B. H., Freude, G., Tempel, J., & Kaluza, S. (2005). *International Congress Series, 1280,* 292–295.

Hechinger, J. (2009, May 20). SAT Coaching Found to Boost Scores – Barely. *The Wall Street Journal*. Retrieved from http://online.wsj.com/news/articles/SB124278685697537839

Heckman, J. J., Stixrud, J., & Urzua, S. (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics , 24*(3), 411 – 482.

Hickey, C. (2008). Physical education, sport and hyper-masculinity in schools. *Sport, Education and Society, 13*(2), 147-161.

Higdem, J. L., Kostal, J. W., Kuncel, N. R., Sackett, P. R., Shen, W., Beatty, A. S., & Kiger, T. B. (2016), The role of socioeconomic status in SAT–freshman grade relationships across gender and racial subgroups. *Educational Measurement: Issues and Practice, 35*, 21–28. doi:10.1111/emip.12103

Holmlund, H., & Silva, O. (2014). Targeting noncognitive skills to improve cognitive outcomes: Evidence from a remedial education intervention. *Journal of Human Capital, 8*(2), 126-160.

Hout, M. (2012). Social and economic returns to college education in the United States. *Annual Review of Sociology, 38*, 379-400.

Hyde, J. S. (2014, January). Gender similarities and differences. *Annual Review of psychology, 65*, 373–398. doi:10.1146/annurev-psych-010213-115057

Ilmarinen, J. (2007). The Work Ability Index. *Occupational Medicine, 57*, 160 doi:10.1093/occmed/kqm008

Ilmarinen, J., Tuomi, K., Seitsamo, J. (2005). New dimensions of work ability. *International Congress Series, 1280,* 3–7. doi:10.1016/j.ics.2005.02.060

Ingalhalikara, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., . . . , Verma, R. (2014, January). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 111*(2), 823-828. doi:10.1073/pnas.1316909110

Irwin, N., (2013, May 7). *Dow 15,000: Here's why caring about it is not as silly as it might seem.* Washington Post. Retrieved from http://www.washingtonpost.com/blogs/wonkblog/wp/2013/05/07/dow-15000-heres-why-caring-about-it-is-not-as-stupid-as-it-might-seem/

Jacob, B. A. (2002, December). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review, 21*(5), 589–598. doi:10.1016/S0272-7757(01)00051-6.

Jaschik, S., (2008). Non-cognitive qualities join the GRE. *Inside Higher Ed*. Retrieved from http://www.insidehighered.com/news/2008/05/22/ets

Johnson, V. E. (1997). An alternative to traditional GPA for evaluating student performance. *Statistical Science, 12*(4), 251-278.

Jonsson, R. (2014). Boys' anti-school culture? Narratives and school practices. *Anthropology & Education Quarterly, 45*(3), 276–292. doi:10.1111/aeq.12068

Juelskjær, M. (2008). Resisting and committing to schooling: Intersections of masculinity and academic position. *International Journal of Qualitative Studies in Education, 21*(1), 49–63.

Kanazawa, S. (2006). IQ and the wealth of states. *Intelligence, 34*, 593–600.

Kappe, R., & van der Flier, H. (2010, February). Using multiple and specific criteria to assess the predictive validity of the Big Five personality factors on academic performance. *Journal of Research in Personality, 44*(1), 142–145

King, K., & Gurian, M. (2006, September). Teaching to the Minds of Boys. *Educational Leadership, 64*(1), 56-61.

Kipnis, D. (1962, February). A noncognitive correlate of performance among lower aptitude men. *Journal of Applied Psychology, 46*(1), 76-80. doi:10.1037/h0040371

Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Muller, K. E. (2008). *Applied regression analysis and other multivariable methods* (4th ed.). Belmont, CA: Thomson Higher Education.

Klevan S., Weinberg, S. L., & Middleton, J. A.. (2016, March). Why the boys are missing: using social capital to explain gender differences in college enrollment for public high school students. *Research in Higher Education, 57*(2), pp 223-257.

Kobrin, J., Patterson, B., Shaw, E. Mattern, K., & Barbuti, S. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Averages.* College Board Research Report No. 2008-5. New York, NY: The College Board.

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence 36*, 153–160. doi:2048/10.1016/j.intell.2007.03.005

Kohlberg, L. (1976). Moral stages and moralization: The cognitive developmental approach. In T. Lickona (Ed.), *Moral development and behavior* (pp. 170–205). New York, NY: Holt, Rinehart & Winston.

Komarraju, M., Ramsey, A., & Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role of academic discipline *Learning and Individual Differences, 24*, 103–109.

Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice, 35*, 11–20. doi:10.1111/emip.12077

Koul, R., Roy, L., & Lerdpornkulrat, T. (2012, July). Motivational goal orientation, perceptions of biology and physics classroom learning environments, and gender. *Learning Environments Research, 15*(2), 217–229. doi:DOI 10.1007/s10984-012-9111-9

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212-218.

Krathwohl, D.R., Bloom, B.S., & Masia, B.B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: The affective domain.* New York: David McKay.

Krohne, H. W., & Hock, M. (2008, December). Cognitive avoidance, positive affect, and gender as predictors of the processing of aversive information. *Journal of Research in Personality, 46*(6), 1572–1584.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., Dunleavy, E., & White, S. (2007). *Literacy in everyday life: Results from the 2003 national assessment of adult literacy.* National Center for Education Statistics, U.S. Department of Education, NCES 2007–480.

Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2011, April). *The role of noncognitive constructs and other background variables in graduate education.* Princeton, NJ: Educational Testing Service.

Lackey, L. W., Lackey, W. J., Grady, H. M., & Davis, M. T. (2003). Efficacy of using a single, non-technical variable to predict the academic success of freshmen engineering students. *Journal of Engineering Education, 92*(1), 41.

Lang, D. M. (2007). Class rank, GPA, and valedictorians: How high schools rank students. *American Secondary Education, 35*(2), 36-48.

Ledsema, R. G., & Obukova, N. (2015, March). Predictors of first-semester success in a small private college. *Atlantic Economic Journal, 43*(1), 161.

Lewin, T. (2014, March 5). A new SAT aims to realign with schoolwork. *New York Times.* Retrieved from http://www.nytimes.com/2014/03/06/education/major-changes-in-sat-announced-by-college-board.html?_r=0

Lind, D. A., Marchal, W. G., Wathen, C. A., & Waite, S. A. (2009). *Basic statistics for business and economics* (3rd Canadian ed.). Columbus, OH: McGraw-Hill. Retrieved from http://highered.mcgraw-hill.com/sites/dl/free/0070951640/354829/lind51640_ch15.pdf

Lipnevich, A. A. & Roberts, R. D. (2012, April). Noncognitive skills in education: Emerging research and applications in a variety of international contexts. *Learning and Individual Differences,* 22 (2), 173–177

Lloyd, J. E. V., Walsh, J., & Manizheh, S. Y. (2005). Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If I'm so smart, why don't I know it? *Canadian Journal of Education, 28*(3), 384-408.

Lounsbury, J. W., Sundstrom, E., Loveland, J. M., & Gibson, L. W. (2003). Intelligence, “Big Five” personality traits, and work drive as predictors of course grade. *Personality and Individual Differences, 35*(6), 1231-1239.

Lovén, J., Herlitz, A., & Rehnman, J. (2011). Women’s own-gender bias in face recognition memory: The role of attention at encoding. *Experimental Psychology, 58*(4), 333-340. doi:10.1027/1618-3169/a000100

MacMillan, D. (2010, June 21). Grockit takes on Kaplan, Princeton Review in $1 billion test-prep market. Bloomberg. Retrieved from http://www.bloomberg.com/news/2010-06-22/grockit-takes-on-kaplan-princeton-review-in-1-billion-test-prep-market.html

Marte, J. (2011, May 5). 10 things test-prep services won't tell you. *The Wall Street Journal*. Retrieved from http://www.marketwatch.com/story/10-things-testprep-services-wont-tell-you-1301943701454?pagenumber=1

Martus, P., Jakob, O., Rose, U., Seibt, R. & Freude, G. (2010). A comparative analysis of the Work Ability Index. *Occupational Medicine, 60*, 517–524. doi:10.1093/occmed/kqq093

Maryland Higher Education Commission (2007). *Meeting agenda memorandum: Retention and graduation rates at public four-year campuses by SAT scores*.

Maryland Higher Education Commission (2011). *Student outcome and achievement report*. Retrieved from mhec.maryland.gov/publications/Documents/Research/AnnualReports/2010SOAR.pdf

Maryland Higher Education Commission (2013). *Retention and Graduation Rates at Public Four-Year Campuses by SAT Scores*. Retrieved from http://mhec.maryland.gov/publications/Documents/Research/AnnualReports/2013RetGrad4yrs.pdf

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689–704

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008, January). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59*, 507-536. doi:10.1146/annurev.psych.59.103006.093646

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and Implications. *Psychological Inquiry, 15*, 197–215.

McClure, J., Meyer, L. H., Garisch, J., Fischer, R., Weir, K. F., & Walkey, F. H. (2011, April). Students' attributions for their best and worst marks: Do they relate to achievement? *Contemporary Educational Psychology, 36*(2), 71–81.

McDaniel, E. D., Halter, A. N., & Hartford, D. (1961). Grade Utility: A New Noncognitive Factor in Academic Prediction. *Educational and psychological measurement, 21*(3), 621

McGanney, M. L., & Ganoo, S. (1995). Noncognitive factors and performance predictions. *Journal of the Association of American Medical Colleges, 70*(1), 1–2

McGregor, H. A., & Elliot, A. J., (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Educational Psychology, 94*(2), 381–395. doi:10.1037//0022-0663.94.2.381

McKelvie, S. J., Standing, L., St. Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society, 31*, 353–421.

McMahon, T. (2011, November 16). What is the inflation adjusted price of corn? Retrieved from http://inflationdata.com/Inflation/Inflation_Articles/Corn_Inflation.asp

Meriac, J. P. (2012, August). Work ethic and academic performance: Predicting citizenship and counterproductive behavior. *Learning and Individual Differences, 22*(4), 549–553

Messick, S. (1979, June). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology, 71*(3), 281-292. doi:10.1037/0022-0663.71.3.281

Moffitt, T., Caspi, A., Harkness, A., & Silva, P. (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful?. *Journal of Child Psychology and Psychiatry, And Allied Disciplines, 34*(4), 455-506.

National Center for Education Statistics. (2007). *Literacy in everyday life: Results from the 2003 national assessment of adult literacy*. Retrieved from http://nces.ed.gov/Pubs2007/2007480_1.pdf

National Center for Education Statistics. (2009a). *ACT score averages and standard deviations, by sex and race/ethnicity, and percentage of ACT test takers, by selected composite score ranges and planned fields of study: Selected years, 1995 through 2009.* Retrieved from http://nces.ed.gov/programs/digest/d09/tables/dt09_147.asp

National Center for Education Statistics. (2009b). *Trends in High School Dropout and Completion Rates in the United States: 1972–2009.* Retrieved from http://nces.ed.gov/pubs2012/2012006.pdf

National Center for Education Statistics. (2011). *Fast Facts – SAT scores.* Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=171

National Center for Education Statistics (2012a). *Average National Assessment of Educational Progress (NAEP) mathematics scale score, by age and selected student characteristics: Selected years, 1973 through 2012.* Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_222.85.asp

National Center for Education Statistics (2012b). *Average National Assessment of Educational Progress (NAEP) reading scale score, by age and selected student characteristics: Selected years, 1971 through 2012.* Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_221.85.asp

National Center for Education Statistics (2012c). *Bachelor's, master's, and doctor's degrees conferred by postsecondary institutions, by sex of student and discipline division: 2011-12.* Retrieved from http://nces.ed.gov/programs/digest/d13/tables/dt13_318.30.asp

National Center for Education Statistics (2012d). *Total fall enrollment in degree-granting postsecondary institutions, by attendance status, sex of student, and control of institution: Selected years, 1947 through 2012.* Retrieved from http://nces.ed.gov/programs/digest/d13/tables/dt13_303.10.asp

National Collegiate Athletic Association (n.d.). *Divisions I and II Initial-Eligibility Requirements.* Retrieved from http://fs.ncaa.org/Docs/eligibility_center/Quick_Reference_Sheet.pdf

Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., Ceci, S. J., …, Urbina, S. (1996, February). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101. doi:10.1037/0003-066X.51.2.77

Neu, T. W. & Weinfeld, R. (2007). *Helping Boys Succeed in School.* Waco, TX: Prufrock Press.

Noble, J. P., & Sawyer, R. L. (2004). Is high school GPA better than admission test scores for predicting academic success in college. *College and University, 79*(4), 17-22.

Nunes, I. L., Costa, A. F., & Puga-Leal, R. (2011). An exploratory study of the Work Ability Index (WAI) and its components in a group of computer workers. *Work, 39*(4), 357-67.

Nyborg, H. (2007). Intelligence, hormones, sex, brain size, and biochemistry: It all needs to have equal causal standing before integration is possible. *Behavioral and Brain Sciences, 30*, 164–166. doi:10.1017/S0140525X07001264

O'Connor, M. C., & Paunonen, S. V. (2007, October). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*(5), 971–990

Olson, A.(2008). Standardized tests. *In 21st century education: A reference handbook.* Retrieved from http://literati.credoreference.com/content/entry/sagetfced/standardized_tests/0

Okun, M. A. (1980). The role of noncognitive factors in the cognitive performance of older adults. *Contemporary Educational Psychology, 5*(4), 321–345.

Pearson. (2012). *The Learning Curve.* Retrieved fromhttp://thelearningcurve.pearson.com/content/download/bankname/components/filename/FINALLearningCurve_Final.pdf

Priest, J. B., Edwards, A. B., Wetchler, J. L., Gillotti, C. M., Cobb, R. A., & Borst, C. (2012). An exploratory evaluation of the cognitive-active gender role identification continuum. *American Journal of Family Therapy, 40*(2), 152-168. doi:10.1080/01926187.2011.601196

Rahman, F., & Masrur, R. (2011, March). Is metacognition a single variable? *International Journal of Business and Social Science, 2*(5), 135-141.

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S. C., Ellis, C., Shakeshaft, … , Price, C. J. (2011, November). Verbal and non-verbal intelligence changes in the teenage brain. *Nature, 479*(7371), 113–116. doi:10.1038/nature10514

Reiss, A. L., Abrams, M. T., Singer, H. S., Ross, J. L., & Denckla, M. B. (1996). Brain development, gender and IQ in children A volumetric imaging study. *Brain, 119*, 1763-1774.

Rindermann, H. (2007). The g-factor of international cognitive ability comparison: The homogeneity of results in PISA, TIMMS, PIRLS and IQ tests across nations. *European Journal of Personality, 21*, 667–706.

Robbins, S. B., Lauver, K., Davis, H. L., Davis, D., Langely, R., & Carlstrom, A., (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261–28. doi:10.1037/0033-2909.130.2.261

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*(1), 103–139.

Ruban, L. M., & McCoach, D. B. (2005, Summer). Gender differences in explaining grades using structural equation modeling. *Review of Higher Education, 28*(4), 475–502.

Sackett, P., Kuncel, N., Arneson, J., Cooper, S., & Waters, S. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135*(1), 1–22. doi:10.1037/a0013978.

Sackett, P. R., & Yang, H. (2000, February). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112-118. doi:10.1037/0021-9010.85.1.112

SAT. (2013a). *SAT Percentile Ranks for 2013 College-Bound Seniors.* Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/SAT-Percentile-Ranks-By-Gender-Ethnicity-2013.pdf

SAT. (2013b). *SAT® Total Group Profile Report.* Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/2013/TotalGroup-2013.pdf

Schapiro, S. R., & Livingston, J. A. (2000, Fall). Dynamic self-regulation: The driving force behind academic achievement. *Innovative Higher Education, 25*(1), 23–35.

Schmidt, J., & Smith, M. C. (2010, April). Girls disengage from high school science. *The Education Digest, 75*(8), 44-45.

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology, 94*(6), 1479–1497. . doi:10.1037/a0016810

Severiens, S., & Ten Dam, G. (1998). Gender and learning: Comparing two theories *Higher Education, 35*, 329–350.

Shah, N. R., & Braverman, E. R. (2012). Measuring adiposity in patients: The utility of Body Mass Index (BMI), percent body fat, and leptin. *PLoS ONE, 7*(4), 1-8. doi:10.1371/journal.pone.0033308

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition, 9*, 313–323. doi:10.1006/ccog.2000.0450

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421-442. doi:10.1037/a0022777

Skelton, C. (1997). Primary Boys and Hegemonic Masculinities. *British Journal of Sociology of Education, 18*(3), 349-369.

Sohn, K. (2010). The role of cognitive and noncognitive skills in overeducation. *Journal of Labor Research, 31*(2), 124-145. doi:10.1007/s12122-010-9082-5

Sparkman, L. A., Maulding, W. S., & Roberts, J. G. (2012, September). Non-cognitive predictors of student success in college. *College Student Journal, 46*(3), 642.

State Higher Education Executive Officers (2012). *State higher education finance FY 2012.*

Täht, K., & Must, O. (2010). Are the links between academic achievement and learning motivation similar in five neighbouring countries? *A Journal of the Humanities and Social Sciences, 14*(3), 271-281.

Tanner, K. D. (2012, Summer). Promoting student metacognition. *CBE- Life Sciences Education, 11*(2), 113-120. doi:10.1187/cbe.12-03-0033

Thomas, L. L., Kuncel, N. R., & Credé, M. (2007, August). Noncognitive variables in college admissions: The case of the non-cognitive questionnaire. *Educational and Psychological Measurement, 67*, 635-657. doi:10.1177/0013164406292074

Toomela, A. (2008). Noncognitive correlates of education. *Learning and Individual Differences, 18*(1), Pages 19-28

Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character.* New York, NY: Mariner Books.

Tracey, T. J. & Sedlacek, W. E. (1984). Noncognitive variables in predicting academic success by race. *Measurement & Evaluation in Guidance, 16*(4), 171–178.

Tross, S. A, Harper, J. P., Osher, L. W, & Kneidinger, L. M. (2000, May/June). Not just the usual cast of characteristics: Using personality to predict college performance and retention. *Journal of College Student Development, 41*(3), 323–334.

Trzaskowski, M., Shakeshaft, N. G., & Plomin, R., (2013, September–October). Intelligence indexes generalist genes for cognitive abilities. *Intelligence, 41*(5), 560–565.

Tucker, J., & Courts, B. (2010). Grade inflation in the college classroom. *Foresight, 12*(1), 45 – 53.

Tyre, P. (2006, January 29). Education: Boys falling behind girls in many areas. *Newsweek*. Retrieved from http://www.newsweek.com/education-boys-falling-behind-girls-many-areas-108593

United Nations Development Programme (2011). *Human Development Report 2011.* Retrieved from http://www.undp.org/content/undp/en/home/librarypage/hdr/ human_developmentreport2011.html

United States Bureau of Labor Statistics (n.d.). *Common Misconceptions about the Consumer Price Index: Questions and answers*. Retrieved from: http://www.bls.gov/cpi/cpiqa.htm

United States Bureau of Labor Statistics (2014). *CPI Detailed Report Data for February 2014.* Retrieved from: http://www.bls.gov/cpi/cpid1402.pdf

United States Department of Energy (1995). *How to measure performance: A handbook of techniques and tools.* Retrieved from http://www.orau.gov/pbm/handbook/handbook_all.pdf

Van de gaer, E., Pustjens, H., Van Damme, J., & De Munter, A. (2006, July). Tracking and the effects of school-related attitudes on the language achievement of males and females. *British Journal of Sociology of Education, 27*(3), 293–309.

Vander Ark, T. (2012, October 31). Non-cognitive skills: Bad name, really important. [Blog Post] Retrieved from http://blogs.edweek.org/edweek/on_innovation/2012/10/non-cognitive_skills_bad_name_really_important.html

Van Houtte, M. (2004). Why males achieve less at school than females: the difference between males' and females' academic culture. *Educational Studies, 30*(2), 159-173.

Vincent-Lancrin, S. (2008). The reversal of gender inequalities in higher education: An on-going trend. *Higher Education to 2030 , 1*. Paris: Organization for Economic Co-operation & Development.

Volwerk, J. J., & Tindal, G. (2012). Documenting student performance: An alternative to the traditional calculation of grade point averages. *Journal of College Admission, 216*, 16-23.

Wan, G. H. (2001). Changes in regional inequality in rural China: Decomposing the Gini index by income sources. *Australian Journal of Agricultural and Resource Economics, 45*(3), 361-381. doi:10.1111/1467-8489.00148

Warne, R. T., Nagaishi, C., Slade, M. K., Hermesmeyer, P., & Peck, E. K. (2014). Comparing weighted and unweighted grade point averages in predicting college success of diverse and low-income college students. *National Association of Secondary School Principals.NASSP Bulletin, 98*(4), 261-279. doi:http://dx.doi.org/0001879977

Warner, R.M. (2013). *Applied statistics: From bivariate through multivariate techniques.* Thousand Oaks, CA. Sage Publications.

Washington College Access Network. (n.d.). Building and measuring meta-cognitive skills at your school or organization. Retrieved from http://www.collegesuccessfoundation.org/2013-statewide-conference/session-3

Weber, P. S., Weber, J. E., Sleeper, B. R., & Schneider, K. L. (2004, February). Self-efficacy toward service, civic participation and the business student: Scale development and validation. *Journal of Business Ethics, 49*(4), 359-369.

Weis, M., Heikamp, T., & Trommsdorff, G. (2013, July 13). Gender differences in school achievement: The role of self-regulation. *Frontiers in Psychology, 4*(442), 1–10.

Willingham, D. (2007, Summer). Critical thinking: Why is it so hard to teach? *American Educator, 31*(2), 8–19.

Willingham, D. (2013, January 21). Measurement of non-cognitive factors. Retrieved from http://www.danielwillingham.com/1/post/2013/01/measurement-of-non-cognitive-factors.html

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*, 1–37. doi:10.1111/j.1745-3984.2002.tb01133.x

Wolters, C. A., & Pintrich, P. R. (1998). Contextual differences in student motivation and self-regulated learning in mathematics, English, and social studies classrooms. *Instructional Science, 26* (1), 27–47.

You, Z. (2010, March). Gender differences in mathematics learning. *School Science And Mathematics, 110*(3), 115–117.

Zimmerman, B. J., & Martinez-Pons, M. (1990, March). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*(1), 51-59. doi:10.1037/0022-0663.82.1.51

Zumbar, M. J, & Blume, L. B. (2008). *Middle childhood development: A contextual approach.* Boston, MA: Pearson.

APPENDIX A

Northeast College IRB Approval

[redacted]

Date: May 12, 2016
TO: Todd Wadsworth
RE: **IRB-16-56**

Approval Date: May 12, 2016
Expiration Date: May 11, 2017

TITLE: ***Creating and Testing a Performance Index to Predict College Performance***

The [redacted] Institutional Review Board (IRB) review of this project is complete and I am pleased to advise that the rights and welfare of the human subjects appear to be adequately protected and methods to obtain informed consent are appropriate. Therefore, the IRB has approved this project.

RENEWALS: IRB approval is valid until the expiration date listed above. Projects continuing beyond this date must be renewed with the renewal form. A maximum of four such expedited renewals are possible. Investigators wishing to continue a project beyond that time need to submit a 5-year application for a complete review.

REVISIONS: The IRB must review any changes in procedures involving human subjects, prior to initiation of the change. If this is done at the time of renewal, please include a revision form with the renewal. To revise an approved protocol at any other time during the year, send your approval and reference the project's IRB number and title. Include in your request a description of the change and any revised instruments, consent forms or advertisements that are applicable.

PROBLEMS/CHANGES: Should either of the following arise during the course of the work, notify the IRB promptly: 1) problems (unexpected side effects, complaints, etc.) involving human subjects, or 2) changes in the research environment or new information indicating greater risk to the human subjects than existed when the protocol was previously reviewed and approved.

If we can be of further assistance, please contact [redacted] or via email [redacted] Please note that all IRB forms are located at [redacted]

Sincerely,

[redacted]

APPENDIX B

Liberty University IRB Approval

Dear Todd Wadsworth,

The Liberty University Institutional Review Board has reviewed your application in accordance with the Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA) regulations and finds your study to be exempt from further IRB review. This means you may begin your research with the data safeguarding methods mentioned in your approved application, and no further IRB oversight is required.

Your study falls under exemption category 46.101(b)(4), which identifies specific situations in which human participants research is exempt from the policy set forth in 45 CFR 46:101(b):

> (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Please retain this letter for your records. Also, if you are conducting research as part of the requirements for a master's thesis or doctoral dissertation, this approval letter should be included as an appendix to your completed thesis or dissertation.

Please note that this exemption only applies to your current research application, and any changes to your protocol must be reported to the Liberty IRB for verification of continued exemption status. You may report these changes by submitting a change in protocol form or a new application to the IRB and referencing the above IRB Exemption number.

If you have any questions about this exemption or need assistance in determining whether possible changes to your protocol would change your exemption status, please email us at irb@liberty.edu.

Sincerely,

G. Michele Baker, MA, CIP
Administrative Chair of Institutional Research
The Graduate School

LIBERTY
UNIVERSITY

Liberty University | Training Champions for Christ since 1971

APPENDIX C

Scatterplot for IPI and FSGPA for RQ 1, RQ 2, and RQ 3

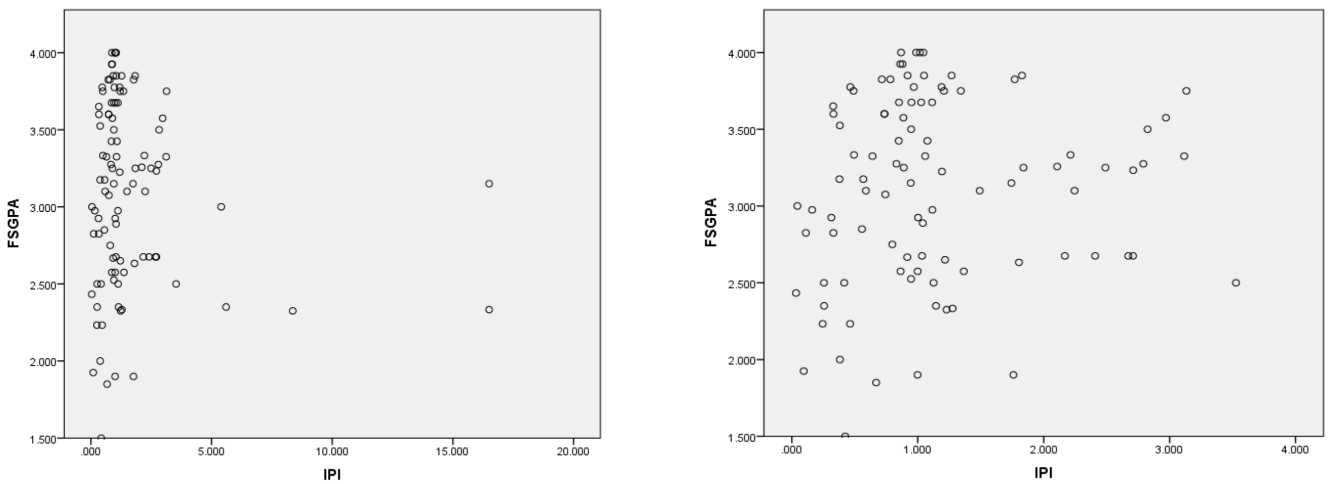

Figure x. Scatter plot for RQ 1 and zoomed for IPI < 4.0 (95 of 100 cases).

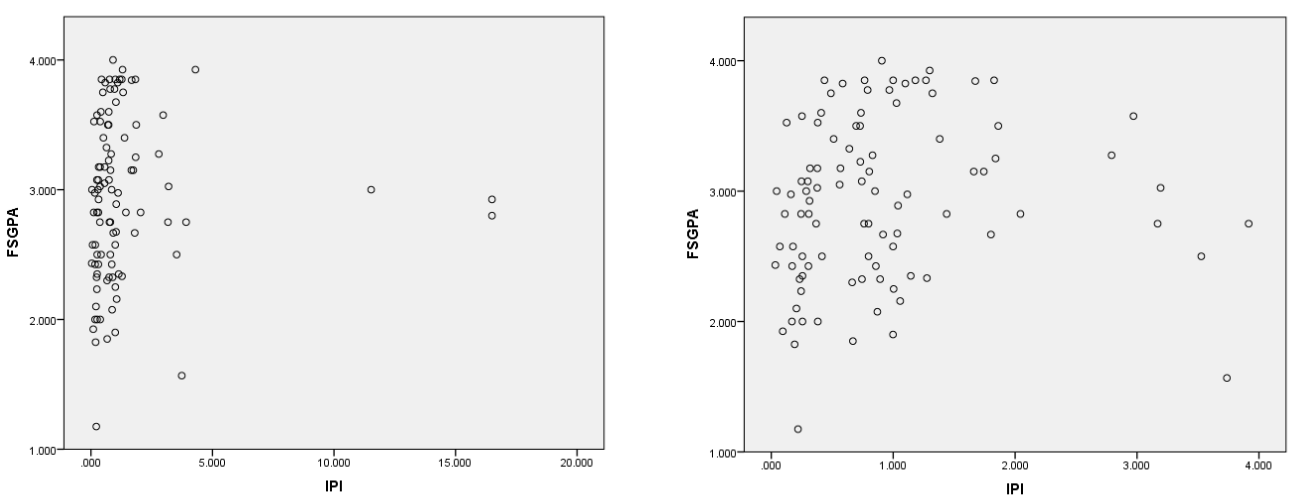

Figure x. Scatter plot for RQ 2 and zoomed for IPI < 4.0 (96 of 100 cases).

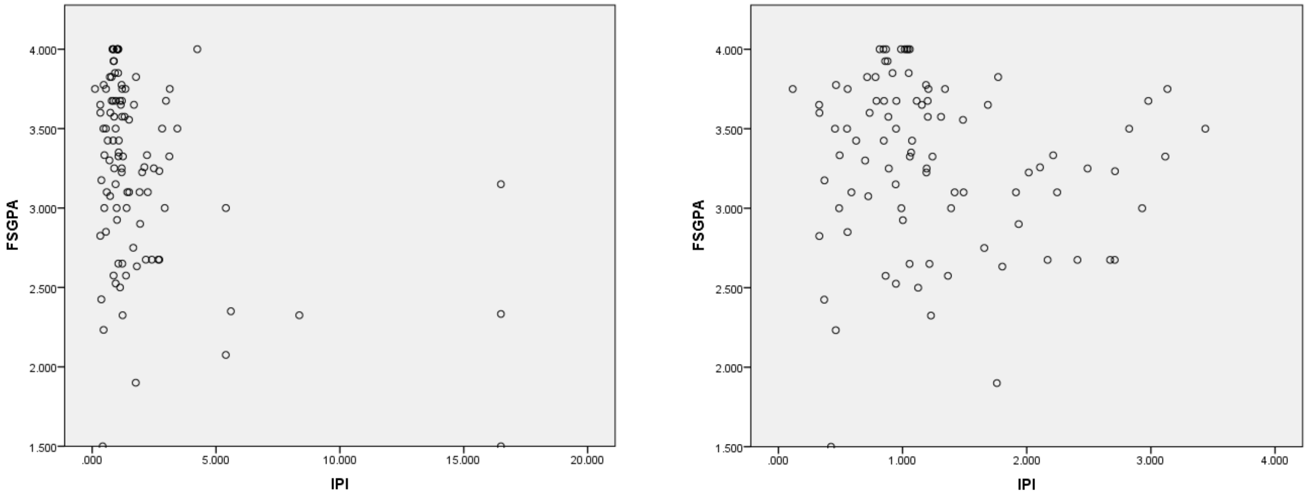

Figure x. Scatter plot for RQ 3 and zoomed for IPI < 4.0 (92 of 100 cases). .