SUBGROUP DIFFERENCES AND PREDICTIVE ABILITY OF PSYCHOMETRIC AND NEUROPSYCHOLOGICAL INTELLIGENCE MEASURES

by

JENNIFER M. FERRETER

A dissertation submitted to the Graduate Faculty of Psychology in partial fulfillment of the
requirements of the degree of Doctor of Philosophy,
The City University of New York

2010

This manuscript has been read and accepted for the
Graduate Faculty in Psychology in satisfaction of the
dissertation requirements for the degree of Doctor of Philosophy.

Charles A. Scherbaum

Date

Chair of Examining Committee

Maureen O'Connor

Date

Executive Officer

Harold Goldstein

Hannah Rothstein

Daniele Artistico

Glenn Albright

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

SUBGROUP DIFFERENCES AND PREDICTIVE ABILITY OF PSYCHOMETRIC AND NEUROPSYCHOLOGICAL INTELLIGENCE MEASURES

by

Jennifer M. Ferreter

Advisor: Professor Charles A. Scherbaum

Researchers recognize that the current models of intelligence are insufficient at making causal connections between the intelligence measure and intelligent behavior. Different approaches to intelligence are under investigation to incorporate within current models of intelligence and include psychometric, neuropsychological, and cultural components. Currently there is a lack of research that incorporates both psychometric and neuropsychological intelligence measures in a predictive model of performance. The purpose of the current study is three-fold. The first objective is to test the predictive relationships of neuropsychological and psychometric intelligence batteries, an alternative psychometric intelligence assessment, and a personality measure in relationship to academic performance. The second objective is to examine racioethnic and gender subgroup mean differences on all predictors of performance. Subgroup mean differences, which can lead to adverse impact, have been found on a variety of verbal and nonverbal intelligence assessments (Hough, Oswald, & Ployhart, 2001). Research has demonstrated that performance differences are often moderated by the type of measure used which also raises concerns about the construct validity of psychometric intelligence assessments. The third objective of the research is to examine the construct validity of neuropsychological intelligence, traditional psychometric intelligence, and alternative psychometric intelligence. There is little empirical evidence which demonstrates that differences in cognitive functioning in the brain result in differences in scores on psychometric assessments. That is, there are few links (i.e., construct validity evidence) connecting cognitive functioning to intelligent performance on psychometric assessments. Hypotheses pertaining to prediction of different measures, subgroup mean differences, and statistical relationships among the intelligence measures were tested. The results indicate that the neuropsychological intelligence battery was the only significant predictor of academic performance. All intelligence measures exhibited subgroup mean differences, however they were smaller compared to what is typically reported in the literature. The Black/African American mean score on the neuropsychological battery was one-third of a standard deviation below the White/Caucasian mean score, and Hispanics demonstrated minimal mean score differences compared to White/Caucasians. Additionally, construct validity evidence emerged for the intelligence measures. A discussion of the findings including their implications is included.

Acknowledgements

This research was shaped and supported by many people. First, I would like to thank my advisor, Charles Scherbaum, for providing insightful feedback on countless drafts and for giving me so many opportunities to work on interesting research. I am enormously grateful for the guidance and supervision throughout our years of working together. I would also like to thank Harold Goldstein and Hannah Rothstein who served on my committee and improved the paper by giving constructive suggestions for clarifying its structure and arguments. This endeavor would have been impossible without the support of my friends and family. My struggles were made more manageable by my wonderful friends who were empathetic to the trials of graduate school. It has been a privilege to work with my manager, Matt Valenti, who was committed to my professional development while being explicitly clear that my dissertation should be my top priority. Thank you for your flexibility and providing the balance I needed to stay productive on all fronts. Most of all, I would like to thank my family whose love and support mean the world to me; especially Dad, Mom, and Mike, who gave me encouragement during every step of this journey, and Mo for providing unconditional comic relief and being the highlight of my day.

> Perseverance is not a long race; it is many short races one after the other.
>
> Walter Elliot, *The Spiritual Life*

Table of Contents

Introduction 1

Chapter 1: Literature Review 8

- Definitions and Operationalizations of Intelligence 8
- Relationship between Cognitive Functioning and Intelligence Scores 13
- Intelligence as a Predictor of Performance 20
  - Using neuropsychology to improve performance prediction. 22
- Subgroup Mean Differences and Adverse Impact on Intelligence Tests 25
  - Using neuropsychology to reduce subgroup mean differences 35
- Validity Concerns for Psychometric Intelligence 37
  - Using neuropsychology to improve construct validity 41
- Summary...45

Chapter 2: Method 47

- Participants 47
- Neuropsychological Measures...47
- Psychometric Measures 51
- Procedure 56

Chapter 3: Results...63

Chapter 4: Discussion...73

Appendix A: Test Attitude Survey-Short 90

Appendix B: Demographic Questionnaire 92

References 103

List of Tables

Table 1: Descriptive Statistics for Predictor and Criterion Variables…………………94

Table 2: Intercorrelations of Predictor and Criterion Measures……….……………....95

Table 3: Tests of Normality for Predictor Variables…………….….……...……….96

Table 4: Regression Models Predicting Academic Performance (GPA)…….…..…..97

Table 5: Subgroup Mean Differences and Effect Sizes (*d*)……………….…..…98

Table 6: Descriptive Statistics for Predictor Variables of White/Caucasian Participants..…………………………………………………………100

Table 7: Percentage of Variance Explained in the Academic Performance Measures by the Cognitive Ability Measures……………….…….……………..…..102

Introduction

Intelligence is a heavily investigated and often debated construct in the social sciences. Although this study examines different approaches to understanding and measuring intelligence, in general, intelligence is how well one understands and processes information. One reason why so many resources are committed to exploring theories of intelligence is because intellectual ability is related to many of life's outcomes (Herrnstein & Murray, 1994; Schmidt & Hunter, 1998). Cattell notes, "There are indeed few measurements in psychology that approach the measurement of intelligence in the frequency with which they are made and the important practical uses to which they are put" (Cattell, 1987, p. 1).

Researchers have studied intelligence (i.e., cognitive ability, general mental ability, *g*) and its impact on various criteria for over a century (e.g., Galton, 1883; Spearman, 1904). The dominant approach to intelligence research is the psychometric approach. The psychometric approach to intelligence is derived from factor analysis and is hierarchical in nature with a general intelligence factor (*g*) at the apex. An individual's intelligence is derived from a test score, which is based on the true score model (i.e., intelligence is synonymous with test performance which is equal to the sum of a true score plus error; Crocker & Algina, 1986). Intelligence theories have been refined over the years, but there is little doubt that intelligence is related to a variety of life outcomes. In a review of the psychometric approach to intelligence, scholars determined that intelligence is related to both academic success and work performance (Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci et al., 1996). Sources estimate both correlates at approximately $r = .50$ (Hunter & Hunter, 1984; Neisser et al., 1996; Schmidt & Hunter, 1998). Regardless of statistical corrections, a relationship of .50 achieved in practice is quite meaningful. Concerning academic success, it means that higher intelligence is related to higher performance in school, which enables admittance to more select colleges, which can then lead to better job offers. Concerning work performance, Herrnstein and Murray (1994) wrote “A smarter employee is, on the average, more proficient employee” (p. 63) and this relationship holds true across professional and skilled occupations, and slightly less for unskilled manual jobs. Once in a job, more efficient work performance can lead to higher supervisor ratings, which can lead to a salary increase and promotions, which begets greater success and higher lifetime earnings. Clearly, performance on psychometric intelligence assessments can greatly impact one’s life trajectory.

There are a number of lingering concerns about research using the psychometric approach, even though it is the dominant approach. The first concern is that job performance is a multiply determined criterion. Therefore, psychometric g alone can only account for a limited amount of variance in the criterion. Psychometric g’s correlation with job performance is $r = .51$, meaning g is able to explain approximately 26% of the variance in job performance (Schmidt & Hunter, 1998). One quarter of variance explained is not trivial, especially not for a single predictor, but it leaves a much larger proportion of the variance unexplained. Even when g is paired with a highly predictive personality trait, conscientiousness, the incremental validity reaches $r = .64$, still leaving 60% of the variance unexplained. These data help to illustrate that a g-only model of intelligence is incomplete. Unique constructs can add meaningful incremental validity to a predictive model. Recent literature (Deary, 2001; Sternberg, 2005) has called for a more comprehensive theory of intelligence which recognizes the contribution of several disciplines (e.g., psychometric, neurobiological, sociological, etc.). Although the call has been made, few researchers have adopted this new research paradigm and traditional ways of thinking about intelligence have not changed markedly.

A second concern is that many tests of *g* produce subgroup mean score differences. Research consistently demonstrates approximately one standard deviation between Black and White test takers on general ability assessments (Hough, Oswald, & Ployhart, 2001; Hunter & Hunter, 1984; Jensen, 1980; Sackett & Wilk, 1994; Williams & Ceci, 1997). Protected classes of applicants (e.g., Black/African Americans, Hispanics, females) score differentially lower on intelligence tests compared to the majority group (e.g., White/Caucasians). The reasons for subgroup mean differences are sharply contested in the intelligence literature (Gould, 1995; Herrnstein & Murray, 1994; Jensen, 1998). Some researchers believe these assessments are simply measuring and reflecting actual differences in general mental ability (Herrnstein & Murray, 1994; Jensen, 1993). Others believe differences in intelligence scores reflect situational causes. As Gould (1995) notes, a score difference "permits no automatic conclusion that truly equal opportunity might not raise the black average enough to equal or surpass the white mean" (p. 13). These two positions will be discussed further in the literature review.

A third concern about the psychometric approach to intelligence is that there are persisting questions regarding the construct validity of *g* (Goldstein, Scherbaum, & Yusko, 2009). A few factors contribute to these validity concerns. First, there is little empirical evidence of a link between cognitive functioning and psychometric test scores. That is, there is little documentation of variations in cognitive functioning that result in variations in performance on a psychometric intelligence test. Psychometric intelligence tests have always relied on correlations with other psychometric intelligence assessments as evidence of validity. The present research will delineate why correlational evidence is incomplete and why the causal link is critical for construct validity. Second, there are many definitions of intelligence. Researchers investigate intelligence in ways that align with their own understandings of the construct. Williams (1996) noted, “[a]ny definition of intelligence carries with it a value judgment about the attributes and performances that are most prized by the society” (p. 506). Many definitions of intelligence invariably lead to many operationalizations. It will be argued later that the lack of consensus on a definition and operationalization is a major weakness of the intelligence construct (in general) and the psychometric intelligence approach (in particular) from a construct validity perspective.

Regardless of the concerns that have been raised, some researchers believe that the all of the questions about psychometric *g* have been resolved and do not need further investigation (Murray & Herrnstein, 1994). Perhaps worse, research on *g* may be avoided due to the contentious nature of its outcomes and implications. Goldstein and colleagues (2009) note “…beliefs of the psychometric approach, especially with regard to the statement that racial differences are inherent in the construct of intelligence, have deterred the field of [Industrial/Organizational] psychology from conducting further research in personnel selection on intelligence and from attempting to develop measures of intelligence that do not produce racial differences” (p.96). The outstanding concerns noted above (i.e., performance prediction, subgroup mean difference, construct validity) indicate that further investigation is needed into the intelligence construct and the methods we use to measure it.

Fortunately, there are a number of approaches other than psychometric for advancing our understanding of intelligence. One that is particularly promising and potentially useful for addressing the concerns about the psychometric approach is the neuropsychological approach to intelligence. This approach to intelligence is not new with its long history in clinical settings, however its application and permeation into the field of Industrial/Organizational psychology is exceptionally limited. The neuropsychological approach to intelligence may be an innovation that is able to push our understanding of intelligence beyond psychometrics. Neuropsychological assessments activate the part of the brain responsible for working memory and executive attention located in the frontal lobes (Conway, Kane, & Engle, 2003; Milner, Petrides, & Smith, 1985). Given the close connection between scores on these assessments and executive functioning in the brain, neuropsychological measures may provide insights into the performance prediction, subgroup mean difference, and construct validity concerns surrounding the psychometric approach.

The purpose of the present research is to test a multifaceted model of intelligence based on the neuropsychological approach as it relates to the prediction of performance in the academic domain, group differences on neuropsychological measures, and relationships between psychometric and neuropsychological measures (i.e., construct validity). Specifically, the current study will measure the predictive relationships of neuropsychological and psychometric intelligence batteries as well as an alternative psychometric intelligence assessment and a personality measure in relationship to academic performance. The following research questions are grouped around the previously identified gaps in the literature. Concerning the performance prediction gap in which psychometric *g* leaves much variance unaccounted for, there are four questions posited:

1. Do neuropsychological intelligence measures positively contribute to predicting performance within a heterogeneous sample?
2. Do other psychometric intelligence predictors and personality factors also positively predict performance?
3. Will the neuropsychological battery have the largest predictive validity coefficient?
4. How relatively important are measures of neuropsychological and psychometric intelligence and personality for predicting performance?

Concerning subgroup mean differences and the possibility of disparate treatment of applicants in a personnel selection context, there are two questions proposed:

5. Do neuropsychological assessments exhibit subgroup mean differences?
6. Will the neuropsychological battery elicit smaller effect sizes compared to other predictors when comparing performance of different racioethnic and gender groups? Or will the alternative psychometric assessment (which previously reported little to no Black/White differences) continue to be the predictor with the smallest effect size?

Finally, concerning the construct validity of psychometric approach to intelligence, two questions are offered:

7. Are psychometric intelligence measures and neuropsychological intelligence measures related to one another (i.e., show construct validity)?
8. Do the validity coefficients vary? Which predictor will be most valid when tested on a racioethnically diverse sample?

These questions were investigated in a non-experimental study within a laboratory setting.

It may be possible that the neuropsychological approach to intelligence is more important in predicting academic success compared to the psychometric approach to intelligence or other predictors of performance (i.e., personality). Previous research (Higgins, Peterson, Pihl, & Lee, 2007) has demonstrated that academic performance was more strongly related to performance on a neuropsychological intelligence battery (r = .37) compared to a psychometric intelligence battery (r = .24) in one study. The results varied for other studies, however when examining intelligence and job performance the reported correlation was r = .52 for all job levels and r = .72 for workers with three or more years of experience. If the present research finds that neuropsychological battery has equal or greater predictive validity and reduced or no group differences compared to psychometric intelligence measures then these findings would be a major contribution to the literature on intelligence testing and the prediction of performance.

Reviews of the intelligence construct and the two main approaches to measuring intelligence are presented in the next chapter. Theoretical and research-based explanations of how the neuropsychological approach to intelligence may help overcome the performance prediction, subgroup mean differences, and construct validity shortcomings of the psychometric approach to intelligence are also included.

Chapter 1: Literature Review

Definitions and Operationalizations of Intelligence

Intelligence is a latent construct, meaning the level of intelligence one possesses cannot be directly observed but is inferred through other measures. The method of inferring intelligence differs between the psychometric and neuropsychological approaches and the disparities have caused confusion and dissatisfaction (Garlick, 2002). The psychometric approach to intelligence research has generated many definitions of intelligence. Jensen's (1980) review of intelligence tests begins with a historical perspective of intelligence definitions produced by the most prominent psychologists from Ebbinghaus ("The power of combination" p. 170) to Humphreys ("…the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies…" p. 170) and Wechsler ("…the capacity…to understand the world about him and his resourcefulness to cope with its challenges" p. 171). In total, Jensen provided more than a dozen definitions of intelligence.

In 1921, and 65 years later in 1986, prominent experts in the field of psychology were asked seemingly simple questions, "What is your conception of 'intelligence'?" and "How is it best measured?" (Sternberg & Detterman, 1986). The 24 experts produced 24 different definitions. Reviewing the definitions, Sternberg and Detterman documented common themes among the descriptions (e.g., addressing how intelligence impacts the individual, his/her behavior, and his/her environment), but each narrative was as unique as its author(s). To be clear, the 1986 experts have not regressed the field's understanding of intelligence; there was very little agreement on the definitions of intelligence in 1921

(Sternberg & Berg, 1986). Thus, concerning how to define the psychometric approach to intelligence, the field seems to be at an impasse.

Intelligence has been defined many ways depending on which aspect of human functioning is emphasized. For the present study, intelligence is defined as information processing, which includes the manipulation of information and not having to rely on prior knowledge. This definition is consistent with the current thinking on intelligence (Neisser et al., 1996) and the inclusion of these elements in the definition is based on established theories of intelligence (e.g., Fagan, 2000) and empirical research (Fagan & Holland, 2002, 2007; Higgins et al., 2007).

Although a working definition of intelligence for the present study was provided, it is important, and perhaps easier, to define what intelligence is not. One thing that most intelligence scholars agree on is that intelligence is different from achievement. Achievement is a measure of how much a person already knows, not a measure of how much a person could learn or is able to learn. At the end of the semester, a professor typically gives a final exam to assess how much a student has learned about a given subject area based on the assigned readings, lectures, and class exercises. The exam is a measure of the student's achievement, not intelligence. The construct of intelligence, the focus of the present research, is entirely different. Intelligence assessments aim to capture intellectual abilities that are not dependent on education or training. Although defining achievement does not necessarily define intelligence, hopefully it is helpful to clarify what is not included in a definition of intelligence.

Before moving forward, it is important to note two additional points. In the psychometric intelligence literature, the terms "intelligence" and "cognitive ability" are used interchangeably. Therefore, to be consistent with the literature, this manuscript will do the same. The terms are considered relatively equivalent for the purpose of the current research. If one has cognitive ability then he or she also possesses intelligence. To address the matter of what is included as a cognitive ability, the research has adopted Carroll’s (1993) interpretation. He notes that, technically, any task can be considered a cognitive task. However, for the purposes of the current research, a cognitive task is limited to “the range of cognitive tasks of those that centrally involve mental function not only in the understanding of the intended end results but also in the performance of the task, most particularly in the *processing of mental information*” (p. 9-10, emphasis original).

Second, “intelligence” as used in this manuscript does not include concepts such as emotional intelligence (EI). Goleman (1995) has made a case for the value of being able to process and regulate one’s own emotions as well as interpret other’s affect, and to make decisions based on this information. Other researchers (Cross & Travaglione, 2003; Gabel, Dolan, & Cerdin, 1995) have provided support that EI is a key predictor of job performance. The author recognizes a place for EI, but the current research is limited to psychometric and neuropsychological approaches to intelligence.

Psychometric approach to intelligence. Briefly, the psychometric approach to intelligence is the most researched intelligence construct and much of the field’s understanding of intelligence is based on psychometric intelligence (Neisser et al., 1996). In one of the first quantitative analyses of intelligence, Spearman (1927) used factor analysis and determined that one dominant factor (*g*) was consistently obtained regardless of the type of intelligence test used in the analysis (i.e., *g* is the shared variance among a variety of cognitive ability assessments). In psychology, and specifically psychometrics, intelligence is *g*. The field of psychology has relied on the model of psychometric intelligence as its primary basis for theory and research involving intelligence. The simplicity and parsimony of a single predictor of intelligence (*g*) on performance is attractive. On the contrary, it may be that *g* is too simplistic to account for the complexities of human intelligence systems.

Some researchers have tried to overcome the concerns of psychometric approach to intelligence by modifying the existing methods of investigation (Fagan & Holland, 2002, 2007; Yusko, Goldstein, Oliver, & Hanges, 2010). This domain of research has investigated what factors elicit subgroup mean score differences in order to eliminate them from the psychometric tests. For example, tasks that are unique, rarely seen in problem solving, or that do not rely on previously learned information tend to show smaller subgroup mean differences. The current study terms these types of measures “alternative psychometric intelligence” assessments because they are still psychometric in nature but do not have the traditional qualities (i.e., relying on previously learned or culture-specific information). One alternative psychometric intelligence assessment is included as a predictor of academic performance in the current study design.

Neuropsychological approach to intelligence. The neuropsychological approach to intelligence has been suggested as an additional or alternative predictor of performance (Deary, 2001; Higgins et al., 2007). One benefit of the neuropsychological approach, unlike the psychometric approach, is that is has a clear definition. Neuropsychology is defined as the study of how the brain relates to behavior (Meier, 1974). The neuropsychological approach to intelligence investigates brain-behavior relationships and relies heavily on biological and psychophysiological measurements. The field of cognitive neuroscience, and specifically clinical neuropsychology, can be associated with unpleasant concepts such as brain damage, lesions, lobotomies, and cerebral dysfunction. Early attempts at mapping cognitive functioning studied individuals with ablations and lesions, and how damage to the brain affected intelligent behavior. However, neuropsychological studies do not necessitate abnormal brain functioning, damage to any part of the brain, or persons who embody a cognitive impairment.

Some neuropsychological research used biological proxies hypothesized to relate to intelligence such as head size and skull size. Attempts to link cognitive functioning to intelligent behavior were somewhat imprecise operationalizations of intelligence such as head size (Rushton, 1994), brain size (McDaniel, 2005), and brain volume (Teasdale & Pakkenberg, 1988). Much of the research had to be conducted posthumously through autopsies to examine the brain and skull. More recently, neurocognitive research has focused on more reliable techniques to measure cognitive functioning. The advent of neural imaging techniques enabled researchers to examine brains of living humans and how brain activity is related to behavior. Three common neuroimaging procedures are electroencephalography (EEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). Neuroimaging can capture activity in the brain including blood flow, glucose consumption, neuronal activity, and can provide images of brain activity. These neuroimaging techniques are used to track neural functioning during spatial and nonspatial tasks including working memory exercises (e.g., recency tasks), word fluency tasks, and association tasks, which are all part of the present study’s design. The next section gives a brief overview neuropsychological approach to intelligence assessments and how technological advances allowed researchers to accurately document the links between brain activation and task performance.

Relationship between Cognitive Functioning and Intelligence Scores.

Conjectures of the brain's complex relationship to intellectual behaviors have been formulating for over 100 years (Horton & Wedding, 1984; Kane & Engle, 2002). Early understandings of neuropsychology were limited to cause-effect and trial and error relationships. An individual consulted a physician with a problem, but often the diagnosis could not come until after the patient was deceased. Patient autopsies revealed lesions, ablations, or other damage to different areas of the brain and each area of the brain was linked to a different ailment. Advancements were slow given that most of the examinations of brain trauma occurred post-mortem.

Many years later, after a period of correlational work that used biological measures of head size, skull size, and brain volume, experimental designs using head trauma patients allowed causal conclusions about the effects of head injuries to be drawn on in vivo patients. Duncan and his colleagues (Duncan, Burgess, & Emslie, 1995) investigated the effects of lesions located in the frontal and posterior regions of the brain (which are divided by the central fissure). They examined eight patients with different brain traumas and found systematic performance deficits based on where the trauma was located. The study included three "frontal" patients and five "posterior" patients and a group of control participants. Seven patients were administered the Wechsler Adult Intelligence Scale (WAIS), and one frontal patient was assessed with the WAIS-Revised (a shorter but comparable version of the WAIS). The WAIS is considered a measure of crystallized intelligence (i.e., the ability to solve familiar problems and recall previous experiences to solve problems). Next, frontal patients were matched with control participants on crystallized intelligence (WAIS) scores, age, sex, and socioeconomic status. All study patients and controls were administered the Culture Fair Test (CFT; Cattell, 1973), a measure of fluid intelligence (i.e., the ability to make inferences and process new information, independent of previously learned knowledge). Although the number of comparisons was small, the results trend to differential outcomes on measures of crystallized intelligence and fluid intelligence, and the differences were mediated by the location of the brain injury.

To interpret the analysis, the comparison metrics were difference scores on the two tests and were calculated by subtracting the Culture Fair score from the WAIS score (e.g., WAIS – Culture Fair; or Crystallized – Fluid). Results for the frontal patients found that the difference scores were much larger ($\bullet_{\text{Frontal}} = 29.6$) compared to the matched or control group ($\bullet_{\text{Control}} = -14.0$), indicating frontal patients outperformed the control group on the WAIS but the control group outperformed frontal patients on the Culture Fair Test. In other words, frontal patients experienced larger deficits in fluid intelligence as measured by the CFT. These differences between the patients' WAIS and CFT scores were considered substantial and significantly different ($p < .01$) based on available reliabilities and the standardized differences (range: $z = 2.7$ to 4.8; Duncan et al., 1995).

Turning to the posterior patients, these five individuals were included in the study to determine if the large deficits in fluid intelligence after brain trauma were typical for all brain injury patients or specific to frontal ablation patients. They found that the location of the brain injury determines the type of cognitive deficient experienced. Posterior patients' WAIS scores, on average, were considerably lower compared to frontal patients (meaning posterior patients experience deficits in crystallized intelligence) but there was no identifiable discrepancy between posterior patients' WAIS and CFT scores.

Duncan and his colleagues' research was beneficial for advancing the neuropsychological literature by showing that intelligence is related to cognitive functioning and identifying specific regions of the brain where crystallized and fluid intelligences are located. However, there are a few notable limitations to this study. First, the sample of eight individuals is small and conclusions should be drawn with some caution about the generalizability of the findings. Another shortcoming of the study design is that there was no control group for the posterior patients. Without a matched control group for the posterior patients it is difficult to discern if posterior patients had deficits in fluid intelligence. Additionally, the researchers selected the frontal patients because of their high WAIS scores, however, the posterior patients' WAIS scores were considerably lower. Therefore, it is difficult to conclude if the lower scores were caused by damage to the posterior regions of the brain or if it was an artifact of the patients selected for the study. Although the case study examines the effects of trauma to the frontal lobes and does less to delineate the effects of posterior brain injuries, the researchers note that "deficits in fluid intelligence are especially marked after frontal lesions" (Duncan et al., 1995, p. 267). Gray and Thompson (2004) conclude, "patients with brain damage provided early data that are still important – *causal evidence* that intelligence behavior depends on the integrity of *specific neural structures*" (p. 472, emphasis added).

Other research documents how patients with damage to the frontal lobes and specifically the prefrontal cortex report great difficulty in mastering everyday cognitive tasks (Lezak, 1983; Shallice & Burgess, 1991). The deficits experiences after prefrontal cortical trauma can include “attention, motor control, spatial orientation, short-term memory, temporal and source memory, metamemory, associate learning, creativity, perseveration, and reasoning” (Kane & Engle, 2002, p. 637). Most or all of these cognitive functions are used in some form during information processing tasks. Taken together, these findings led researchers to conclude that there is a causal relationship between brain structures and intelligence behavior. This conclusion is especially meaningful to intelligence researchers because there is limited research evidence of causal relationships in the psychometric intelligence domain. Providing empirical evidence for a causal relationship is one important step in establishing construct validity of any approach to intelligence. Neuroimaging techniques provided additional evidence for construct validity and further strengthened our understanding of the connection between neural functioning and intelligent behavior.

The introduction of neural imaging techniques helped researchers examine how the brain relates to both input stimuli and behavioral output. The three most common techniques are EEG, PET, and fMRI. EEG measures a brain’s electrical activity through sensors placed on the scalp. P300, an event-related potential occurring 300 ms into an EEG recording, is positively correlated with intelligence but the relationship is inconsistent (Cianciolo & Sternberg, 2004). Another neuroimaging technique, PET, is often combined with computed (axial) tomography (i.e., CAT or CT scans) to produce an image of a selected plane of the brain. The images record hemodynamic changes (i.e., blood flow) and glucose consumption while individuals are engaged in activities. Glucose consumption is considered one measure of neural efficiency. The results have been mixed, but there is increasing evidence that individuals with higher cognitive ability exhibit less glucose consumption, indicating their brains are more efficient and use fewer resources for intellectual tasks (Haier, 2003; Haier, Siegel, MacLachlan, Soderling, Lottenberg, & Buchsbaum, 1992). Another technique used to measure blood flow and neural activity is functional magnetic resonance imaging. fMRI is similar to PET however fMRIs do not require the consumption of radioactive isotopes and can capture multiple images in a shorter amount of time. All three neuroimaging techniques are used in a variety of studies to examine how neural processes directly relate to intellectual behavior and performance on cognitive tasks.

Recently, neuroimaging has been combined with executive functioning tasks such as those used in this study to determine what areas of the prefrontal cortex in the brain are activated during information processing. Prabhakaran and colleagues (Prabhakaran, Narayanan, Zhao, & Gabrieli, 2000) used fMRI to observe and locate areas of the brain that are activated for different working memory tasks. The two different types of tasks were integrated (i.e., combining multiple pieces of information, such as verbal and spatial) and unintegrated (i.e., verbal and spatial information were kept separate). Visual stimuli for the study were displays with four letters and four spatial locations (e.g., locations were specified in the form of parenthetical brackets) presented for 2 seconds. For integrated conditions the letters and locations were presented together, that is, a letter is inside parentheses and the parentheses specify the letter’s locations (e.g., (M)). Unintegrated information presented four letters in a straight line with none of the letters enclosed within the parentheses. Their hypothesis, that the frontal cortex has a specific function in maintaining integrated information within working memory, was supported. Integrated stimuli produced higher levels of activity in the right prefrontal cortex compared to unintegrated stimuli. Unintegrated stimuli produced higher levels of activity in multiple posterior regions of the brain, including bilateral parietal (which lies behind the central fissure), temporal (side of the brain), and cerebella (i.e., cerebellum, located under the parietal and occipital lobes) regions. Similar results were found in literature reviews pertaining to working memory and general intelligence (Conway et al., 2003; Kane & Engle, 2002). The tasks in Prabhakaran's research are similar to those used in the current study. The neuropsychological battery contains multiple tasks that would be considered integrated. For example, one task starts with 12 words that are presented in a spatial array. The objective is to click on each word only once. Once a word has been selected the spatial array changes. The need to remember 12 words and the approximate location they were displayed uses the working memory and attention functions of the brain.

As the literature advances, our understanding of the relationship between neural functioning and intelligent behavior is more detailed and pinpoints to even more specific regions of the brain. In a review of working memory capacity and reasoning ability literature, the importance of a specific area of the prefrontal cortex, called the dorsolateral prefrontal cortex (DLPFC) is delineated (Conway et al., 2003). The DLPFC is associated with working memory, executive control, and *g* (however, they define *g* as general reasoning ability, not general intelligence, which more accurately reflects fluid intelligence). The DLPFC is especially important for working memory tasks that require storage plus additional processing compared to storage only tasks. The difference between the storage and storage-plus tasks is substantial, even producing different neuroimaging patterns. Thus, it is the integration or additional processing component of working memory which incites the prefrontal cortex, specifically the DLPFC. A different task in the current study's neuropsychological battery is one in which the participants need to learn and recall different pairings of words (i.e., 5 cue words paired with 5 target words). The visual display of the target words also changes after each selection so the participant cannot rely on the spatial array to remember word pairings. Again, to successfully complete this task, the participants rely heavily on working memory and executive attention.

Summary. The key concepts to review prior to moving to the concerns of the present study are the intelligence definition, an understanding of the psychometric and neuropsychological approaches to intelligence, and the differences in available evidence for these two approaches. As a reminder, the current study defines intelligence as information processing which does not rely on previously learned knowledge or culture-specific experiences. There are two approaches used in the current study: psychometric, which is the approach that is used in the majority of intelligence research, and neuropsychological, which is an emerging approach that has empirical links to the brain, especially when neuroimaging techniques are used. Both approaches to intelligence research will be explained further in the following sections of the paper. The remainder of the literature review will focus on the previously identified shortcomings of the current literature as it pertains to intelligence research. Specifically, the concerns of psychometric intelligence are reviewed and explanations of how the neuropsychological approach to intelligence may provide insights into performance prediction, subgroup mean differences, and construct validity are provided.

Intelligence as a Predictor of Performance

General mental ability is a strong predictor of academic success and job performance (Neisser et al., 1996). Schmidt and Hunter (1998) call *g* the single best predictor of performance. However, some argue that a single predictor may be insufficient given the complex nature of human intelligence. According to meta-analytic findings based on data gathered over 80 years, general mental ability alone explained 26% of the variance in performance prediction, leaving nearly 75% unexplained (Schmidt & Hunter, 1998). Campbell (1990) offered two suggestions for improving prediction. First, he suggested developing a variety of measures to assess the target construct. As Campbell (1990) notes, "[s]urely a brief paper-and-pencil measure of such a broad construct as general cognitive ability can't be the best that is possible" (p. 690). Moving beyond tests of *g*, especially tests which rely on previously learned information, could improve prediction. Campbell goes on: "[t]he second strategy is for several predictors to be combined into a composite" (p. 690). In a modern model of intelligence, it is only when predictors are combined can more variance in the performance criterion be explained. The current research puts both suggestions into practice by using a neuropsychological intelligence battery which is quite different from traditional psychometric paper/pencil intelligence batteries. Additionally, the neuropsychological battery is combined with other intelligence assessments and a personality measure which have shown to positively contribute to incremental validity (Schmidt & Hunter, 1998).

The position of the current research is that psychometric intelligence theories alone are insufficient at explaining how intelligence is causally related to performance outcomes. These limitations necessitate theorists to expand their current conceptualizations of intelligence to using systems theories (Cianciolo & Sternberg, 2004; Deary, 2001). Cianciolo and Sternberg (2004) note that no single theory can adequately describe intelligence and that multiple perspectives are needed. Previous research investigated *g* in conjunction with a variety of intellectual operations including working memory (Conway et al., 2003), perceptual speed (Mount, Oh, & Burns, 2008), general frontal lobe functioning (Duncan, 1995), neural plasticity (Garlick, 2002), and cross-cultural differences (Helms-Lorenz, Van de Vijver, & Poortinga, 2003). All of these concepts relate to general intelligence in varying degrees, but none is accepted as the unitary cause of *g* (Van der Maas, Dolan, Grasman, Wicherts, Huizenga, & Raijmakers, 2006). Without forsaking parsimony in intelligence models and theories, researchers are cognizant that the complexities of the intelligence construct necessitate the fusion of multiple research programs to understand the construct at different levels of human systems. Using a multi-perspective approach to intelligence is required to advance the understanding of one of psychology’s most nebulous constructs.

There is support in the literature for trying to move the discussion away from a *g*-only model of intelligence (Goldstein, Zedeck, & Goldstein, 2002; Murphy, 1996). Goldstein and colleagues (2002) argue that the single-solution of *g* has rendered the research surrounding performance prediction into submission. They argue that “*g* should be viewed as a starting point rather than an ending point” (p. 124). The current study aims to go beyond a *g*-only psychometric approach for intelligence to predict performance.

Much of the variance is unaccounted for when using only psychometric measures of intelligence which necessitates looking to other approaches to intelligence or different constructs (e.g., personality) to add incremental understanding to such a complex process as human intelligence.

Using neuropsychology to improve performance prediction. Neuropsychological intelligence may be the frontier in intelligence testing however much more research is needed. Blending different approaches to intelligence and comparing each approach's predictive ability is a relatively new practice in the intelligence research domain. One innovative study (Higgins et al., 2007) looked at the predictive validities of neuropsychological and psychometric intelligence, among other variables, for academic and workplace performance. As previously noted in the Introduction, they found that a neuropsychological intelligence battery was able to predict academic performance and job performance (in the form of supervisory ratings). In one study, the neuropsychological battery was better able to predict academic performance than a traditional psychometric intelligence battery. Their research was one of the first known studies examining both neuropsychological intelligence and psychometric intelligence; however the study design contained many limitations. One study examined the predictive ability of both intelligence batteries using a female-only sample enrolled at Harvard University. Such limited sampling reduces the generalizability of their findings. The second study used a sample that is plausibly more representative of the general population (male and female students from University of Toronto in Ontario, Canada) however some students were monetarily compensated and some were given course credit.

The current study will draw from a racioethnically diverse pool of participants who will be equally compensated for their time.

One contribution of the current study is to test the predictive relationships of neuropsychological and psychometric intelligence batteries to determine if the results from Higgins and colleagues (2007) can be replicated with a diverse sample. It is important for the purpose of external validity that the sample drawn for the study is racioethnically diverse. To date, the Higgins study is the only known account of comparing neuropsychological and psychometric intelligence batteries and given the limited amount of research that focuses on these types of variables, there is value added in replicating previous research to see if the predictive relationships hold. Based on previous findings, it is hypothesized that:

> *Hypothesis 1a*: The neuropsychological intelligence battery will positively predict academic performance (GPA).
>
> *Hypothesis 1b*: The traditional psychometric intelligence battery will positively predict academic performance (GPA).

In addition to the two intelligence batteries, an alternative psychometric intelligence assessment is included in the study. Previous research in a laboratory setting (Ferreter, Goldstein, Scherbaum, Yusko, & Jun, 2008) found that the alternative psychometric intelligence assessment was a stronger predictor of academic performance as measured by GPA ($r = .44$) compared to a more traditional psychometric intelligence assessment, the Wonderlic Personnel Test ($r = .33$). Thus, it is hypothesized that:

> *Hypothesis 1c*: The alternative psychometric intelligence assessment will positively predict academic performance (GPA).

In addition to assessments of intelligence, non-cognitive constructs have also positively predicted performance. The “Big Five” personality factors (conscientiousness, openness, extraversion, emotional stability/neuroticism, and agreeableness) each relate to job performance in varying degrees (Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006). Of the five personality dimensions, conscientiousness is the strongest predictor of performance. A meta-analysis reported the estimated true validity of conscientiousness at $r = .20$, and it also predicted sales and customer service job performance at .26 and .25, respectively (Hurtz & Donovan, 2000). When *g* was paired with conscientiousness the multiple *R* was .60, which was one of the highest pairings tested meta-analytically (Schmidt & Hunter, 1998). There is value in testing conscientiousness as its own predictor of academic success and pairing it with other predictors of academic performance. Therefore, it is hypothesized that:

> *Hypothesis 2a*: Conscientiousness will positively predict academic performance (GPA)
>
> *Hypothesis 2b*: Conscientiousness will add incremental validity when paired with the neuropsychological intelligence battery to predict academic performance (GPA)
>
> *Hypothesis 2c*: Conscientiousness will add incremental validity when paired with the psychometric intelligence battery to predict academic performance (GPA)

There is an invested interest to discover the predictive validities of the neuropsychological and psychometric intelligence batteries, especially to know which battery has the larger predictive validity coefficient. Previous research demonstrated mixed results concerning which intelligence battery (neuropsychological or psychometric) has a higher predictive validity coefficient (Higgins et al., 2007), thus it is difficult to hypothesize the expected findings of the current study. However, in the same research the neuropsychological battery was a better predictor of work performance. Therefore it is hypothesized that:

> *Hypothesis 3a*: The neuropsychological intelligence battery will have a larger predictive validity coefficient for academic performance (GPA) compared to the traditional psychometric intelligence battery.

The alternative psychometric intelligence assessment has not been used in conjunction with a neuropsychological intelligence assessment before therefore it is uncertain which measure will have a larger predictive validity coefficient. However, if the separate literatures are an indication of the magnitude of expected coefficients, then it is hypothesized that:

> *Hypothesis 3b*: The neuropsychological intelligence battery will have a larger predictive validity coefficient for academic performance (GPA) compared to the alternative psychometric intelligence assessment.

Higgins and colleagues' (2007) research on the predictive validity of neuropsychological and psychometric intelligence batteries is encouraging, however these findings have yet to be replicated. Furthermore, no data on subgroup mean differences were reported. No known research has reported potential subgroup mean differences or any other negative effects on score differences from using neuropsychological assessments. Recommending the use of a neuropsychological assessment without reporting potential subgroup differences is short of due diligence. Consequently, an explanation of subgroup mean differences, including potential causes and how neuropsychological intelligence tests may reduce group differences is presented in the next section.

Subgroup Mean Differences and Adverse Impact on Intelligence Tests

Through the years, group differences research "has had a disproportionately large (and strongly negative) impact on the public perception of intelligence research" (Gray & Thompson, 2004, p. 479). It is well documented that there is a large gap in performance scores between White and Black test takers (Hough et al., 2001; Roth, Bevier, Bobko, Switzer, & Tyler, 2001) It is also well established that there is much greater variance in intelligence within a racioethnic group compared to across groups (Jensen, 1980; Loehlin, 2000). However, the focus of much research is on the between-group differences (e.g., Black/African American and White/Caucasian, Hispanic and White/Caucasian, male and female). The evidence supporting differences in average intelligence test scores of racioethnic groups has been used to espouse past racial discrimination (Neisser et al., 1996). Throughout this section, group differences will be reviewed based on observed differences in mean scores on various cognitive ability measures. These observed differences in performance, called effect sizes, can be quantified using Cohen's (1988) *d* statistic in order to put the differences onto a common metric. If mean score differences are large enough, it could result in adverse impact in a selection context.

The following sections give an overview of the potential causes of subgroup differences on cognitive ability assessments, examining Black/African American and White/Caucasian racioethnic groups. All adverse impact is serious, but historically the largest and most consistent differences on cognitive assessments are Black/White differences. The literature is rich with explanations for these differences, with the most prevalent reasons being attributed to genetic and environmental causes (Herrnstein & Murray, 1994; Nisbett, 2005; Rushton & Jensen, 2005). Genetic and environmental antecedents are considered contradictory in the literature however both explanations are offered. Additionally, an alternative explanation to Black/White differences that circumvents the nature versus nurture debate is presented. The alternative explanation shows compelling evidence that the way in which cognitive ability is assessed may be substantially contributing to the large disparities in test scores between Black and White test takers. Some alternative assessment methods and the implications for psychometric *g* are further discussed below.

Genetic differences. Knowing how much genetic material people share as well as knowing their intelligence scores allows the calculation of the heritability estimate. The heritability estimate is the degree to which genetics contribute to the variation observed in a trait (Herrnstein & Murray, 1994). Similar to correlations, which have a theoretical range of -1.0 to +1.0, the heritability estimate has a theoretical range of 0 to +1.0, and like correlations the theoretical values of 0 and +1.0 are never obtained in practice. Low heritability estimates (i.e., values closer to 0) indicate that much of the variation in one group's intelligence scores is due to environmental causes. That is, genetics have little or no impact on the differences in intelligence scores across individuals. High heritability estimates (i.e., values closer to 1) indicate that much of the variation in a group's intelligence scores is due to genetics.

Heritability estimates for psychometric intelligence vary depending on the source. Some posit estimates between .40 and .80 (Herrnstein & Murray, 1994, p. 105) with more focused ranges for particular populations (e.g., Grigorenko, 2000, estimated heritability between .50 to .70 for White-middle class preadolescents and teenagers). It is important to note that heritability estimates calculate the approximate importance of genetic-environmental factors influencing a group of people or a population, not an individual. Moreover, heritability estimates do not indicate that genes cause or inhibit intelligence differences between groups. There are some "nature" proponents (Herrnstein & Murray, 1994; Jensen, 1998; Rushton & Jensen, 2005) who assert that differences in intelligence are due to different genetic compositions of Black and White test takers. However, the APA task force did not find support for this assertion (Neisser et al., 1996).

Environment. Scholars who support that mean group differences are caused by factors other than inherited genetic differences offer numerous alternative explanations. One researcher alone has proposed over 120 possible reasons for Black/White mean score differences in cognitive ability tests none of which are inherited (Wiesen, 2007). Other research also supports an environmental basis for intelligence. One study examined German children born of Black and White American World War II soldiers living in the region during the post-1945 occupation (Eyferth, 1961 cited in Nisbett, 2005). Eyferth reported that children fathered by Black GIs had the nearly the same IQ as children fathered by White GIs (i.e., one-half IQ point difference). Nisbett (2005) also pointed out that environmental effects were probably worse for Black children, implying that they had to overcome more than White children to have equivalent intelligence scores.

Decades later, another study examined the intelligence scores of Black children adopted by White and Black middle-class families. Moore (1986) observed score differences for 46 children, equally divided between traditionally adopted (i.e., Black child with Black family) and transracially adopted (i.e., Black child with White family) groups. Performance on the full scale Wechsler Intelligence Scale for Children (WISC) for the transracially adopted Black children (into White families) was within the average range of the assessment. The traditionally adopted Black children (into Black families) scored 12.5 points lower, almost a full standard deviation. These results were similar to a previous study on transracially and traditionally adopted children (Scarr & Weinberg, 1976). Moore concluded that the differences are not simply due to differences in genetics but rather they are “significantly influenced by the ethnicity of the rearing environment” (p. 321).

Currently, the cause of performance differences cannot be determined with certainty (Neisser et al., 1996). However, even after a critical examination of the literature while weighing the evidence presented for genetics and environmental antecedents, neither argument provides a satisfactory explanation of Black/White mean score differences. Researchers are forced to look at alternative explanations for group differences. Conclusions drawn from a meta-analysis examining the determinants of differential subgroup scores are beneficial for further understanding how factors other than genetics or environment can impact performance on cognitive ability assessments. Hough and her colleagues (2001) examined group differences across several predictor domains used in personnel selection, including cognitive ability. They reported varying effect sizes (i.e., *d*) for different measures of intelligence within and across racioethnic groups. Measures of general intelligence had the largest effect sizes along with other measures of crystallized intelligence. Measures of fluid intelligence had, on average, smaller effect sizes. As for lessons learned from their analysis they note that “…the setting, sample, and the *construct* can all, either individually or in combination, moderate the magnitude of the differences between the groups…” (p. 160, emphasis added). Given that the type of test (i.e., the subdimension of intelligence that is assessed) determines the performance differential – not differences in the test takers’ genetics – is one of the strongest counter arguments of the nature debate presented in the literature thus far. An acknowledgement that a test may be a determinant (or *the* determinant) of performance on a cognitive ability assessment is relatively new idea in intelligence research. The next section describes a small portion of innovative research that is currently changing the traditional ways in which scholars assess intelligence.

Measurement of the intelligence construct. The nature versus nurture debate is steeped in controversy. The controversy is, in part, that some arguments are not always based on scientific research (Neisser et al., 1996) and partly because some psychometric intelligence researchers interpret subgroup mean differences "as reflecting reality rather than representing bias or measurement problems and thus those subscribing to this perspective conclude that White test takers actually possess greater intelligence than Blacks" (Goldstein et al., 2009, p. 118-119). Given that there are a number of questions still not addressed in the current literature concerning sources of group differences on intelligence tests, an examination of other explanations (i.e., measurement problems) is warranted.

The research presented next focuses on aspects of cognitive ability assessments that are identified as both contributing to score differences (i.e., word difficulty, typical word use of different racioethnic groups) and elements that successfully decreased the gap in scores between Black and White test takers. Current research programs are exploring the impact of personal experiences and how an individual's background can influence one's knowledge, including the level of specificity of knowledge. These efforts have led to encouraging findings when trying to reduce Black/White differences.

In a group of studies by Fagan and Holland (2007), participants were given a problem solving task for which some of the items required general knowledge (e.g., information that most or all participants have access to: what does "an apple a day keeps the doctor away" mean?) and some problems which required specific knowledge (e.g., exposure to particular knowledge or facts in their past experiences: what city is “home of the bean and cod”?). Across studies, they report that White test takers scored higher than Black test takers on the tests. However, they also found an interaction between race and type of knowledge needed for comprehension (i.e., general versus specific). White test takers performed significantly better compared to Black test takers on the problems that required specific information, but there were no differences for items that required general information. Based on five studies, the authors concluded that Black and White test takers “process information equally well (i.e., they are *equally intelligent*) when the information to be processed has been made *equally available* to all groups” (p. 322; emphasis added).

Sternberg and his colleagues (Sternberg, 1981, 2006) have created more broad and heterogeneous intelligence measures by expanding their operationalization of intelligence. They have also devoted much of their research efforts to developing alternative predictors of “successful” intelligence, including alternative predictors of college performance (Sternberg, 2006). Sternberg’s research often uses entrenched tasks (i.e., which require typical or familiar problem solving techniques) and nonentrenched tasks (i.e., which are novel problems to the participant and he or she must use atypical strategies to resolve the issue). One example of a nonentrenched task is to create rules for fictional objects or characters (e.g., a *plin*) and then ask participants questions about the object as the rules pertain to it. The nonentrenched task performance correlated with performance on psychometric tests at the same strength as psychometric tests correlate with each other (Sternberg, 1981, p. 14). Additionally, nonentrenchment was proposed as an important factor for understanding individual differences in cognitive ability test performance. Sternberg’s approach of novel, creative, and nonentrenched tasks has positive outcomes when the goal is to reduce adverse impact on minorities without losing predictive validity of the assessment (termed the “diversity-validity dilemma” by Pyburn, Ployhart, & Kravitz, 2008).

Ferreter and colleagues (2008) synthesized these previous successful attempts at reducing Black/White differences into one assessment called the Siena Reasoning Test for the purpose of ameliorating subgroup mean differences and maintaining validity. This alternative psychometric intelligence assessment included novel and unique tasks (i.e., using nonsense words), general knowledge (i.e., easily identifiable vocabulary words such as *lamp*) and provided all participants equal access to the information needed to solve the verbal assessment. In a study comparing multiple cognitive ability tests (Ferreter et al., 2008) the differences in effect sizes for the Ravens Progressive Matrices (RPM), the Wonderlic Personnel Test (WPT), and the Siena Reasoning Test (SRT) were substantial. The authors reported the smallest effect size for the SRT (d = -.06 slightly favoring Black/African Americans), followed by the RPM (d = .59, favoring White/Caucasians), and then the WPT (d = .71, favoring White/Caucasians). Performance data were collected in the form of GPA and SAT scores for verbal and quantitative abilities. Furthermore, the SRT explained more variance in GPA and SAT verbal scores compared to the RPM and the WPT.

Freedle (1986) and his colleague Kostin (Freedle & Kostin, 1988, 1990, 1997) have repeatedly found Black/White performance differences on verbal items when analyzing different scholastic exams. However, instead of examining group differences using effect sizes, they used a statistical technique called Differential Item Functioning

(DIF). DIF analyses compare two groups in which group members are matched based on a specified variable. DIF occurs for an item when respondents from two disparate groups who have been matched on ability have different probabilities of correctly answering an item. If an item demonstrates DIF, it indicates the two groups performed differently as a function of a secondary characteristic such as racioethnic group or culture (Camilli & Shepard, 1994). DIF coefficients can either be positive or negative. In Freedle's work, positive DIF indicates that the item favors Black test takers, that is, Blacks are more likely to answer the item correctly compared to their matched White counterparts. On the contrary, negative DIF designates that the item favors White test takers.

Freedle (1986) examined how the difficulty of an item impacts DIF. Freedle specifically examined the performance on verbal items from the SAT and GRE. He reported positive DIF for more difficult SAT analogy items, meaning Black test takers were more likely to correctly answer a harder question than White test takers or the groups performed equally (i.e., no DIF). The opposite was true for easier items. White test takers were systematically more likely to answer easier analogy items correctly. Other research has replicated these results of Black test takers performing differentially better compared to White test takers on difficult items. Freedle and Kostin (1990) found related results using GRE verbal items instead of SAT items. Similarly, Raju, Drasgow, and Slinde (1993) investigated 45 verbal items and divided the items into two equal groups based on median difficulty. The authors reported that 16 of the 22 difficult items (73%) favored Black test takers and 14 of the 22 easier items (64%) favored White test takers. Scherbaum and Goldstein (2008) report comparable findings using item response theory methods to examine DIF. Although the findings from all studies comparing DIF and item difficulty are interesting, it is even more compelling to note that three of these studies reported no DIF for gender groups (i.e., the other study did not test DIF for gender groups). Freedle and others (e.g., Sternberg, 1984) are cognizant of the role that culture and personal experiences can play which shapes an individual's perceptions (Freedle & Kostin, 1997, termed it the "cultural familiarity hypothesis", p. 421). Therefore, the latent mechanism of the verbal ability tests that educes performance differences for Black and White test takers and not male and female test takers is racioethnic in nature and it is drawing on something very specific about the examinee's cultural context.

To review, it is generally accepted that White test takers outscore Black test takers on cognitive ability assessments by approximately one standard deviation (Hough et al., 2001). There are many different explanations for what causes subgroup mean differences, however one explanation is central to the present study: when the intelligence construct is measured different ways, the subgroup mean differences also vary (Hough et al., 2001; Roth et al., 2001). In other words, the measure matters. The largest effect sizes in Black/White performance come from measures of general intelligence, that is, assessments which require the participant to use knowledge that was previously learned either from one's education or upbringing. Cognitive ability measures that necessitate the processing of information that was not learned prior to starting the experiment produced the smallest Black/White effect sizes (Fagan & Holland, 2007; Ferreter et al., 2008). Additionally, prior research using an alternative psychometric intelligence assessment was successful in eliminating Black/White mean score differences using nonentrenched test of analogies (Ferreter et al., 2008). Analogy test items that contained nonsense words and words that all test takers would have learned prior to the experiment produced no mean score differences (*d* = -.06, with a negligible advantage for Black/African Americans).

Using neuropsychology to reduce subgroup mean differences. A growing body of research has identified factors that lead to Black/White mean score differences. An assessment that is novel and does not require previously learned knowledge (i.e., all test takers have access to information needed to problem solve) typically results in reduced or no subgroup mean differences. The neuropsychological intelligence battery used in the current study meets these conditions. Neuropsychological assessments, specifically those assessments used in the present study, focus on conditional association, working memory, and word generation tasks. Conditional tasks require the participant to learn behavioral rules (e.g., If Prompt A, then Respond with Behavior Y; if Prompt B then Respond with Behavior Z). Working memory, the "ability to integrate diverse forms of information in current thought" (Prabhakaran et al., 2000, p. 85) and executive attention, "a capability whereby memory representations are maintained in a highly active state in the presence of interference" (Kane & Engle, 2002, p. 638), are both components of problem solving and reasoning associated with fluid intelligence. Both of these types of tests have reportedly lower subgroup mean differences. Hough and colleagues (2001) report that tests of working memory typically have a Black/White effect size of *d* = .50 (effect sizes were not reported for a Hispanic/White comparison). Ployhart and Holtz (2008) also note that using narrower measures of ability tend to have smaller group differences compared to general ability measures and recommend it as a strategy for overcoming the diversity-validity dilemma. An example of a working memory task from the neuropsychological intelligence battery is the self-ordered pointing task. For this task, there is a presentation of stimuli and the objective is to point to (or click on if computerized) each stimulus object only once. The stimuli will change locations after each trial thereby removing the aid of spatial cues. Word fluency tasks are also categorized as neuropsychological tasks. Word fluency tasks require the participant to generate words that meet particular criteria (i.e., words that start with a certain letter or words that rhyme with a target word).

All of these tasks should reduce subgroup mean differences. The tasks do not require any previously learned information in order to perform well. Additionally, all participants have equal access to the information, which should also reduce any differences. The present study will use a heterogeneous sample to test comparisons that were not available in previous research. One unique contribution of the current study is that both racioethnic (i.e., Black, Hispanic, and White) and gender subgroup differences, if any, will be reported. The neuropsychological intelligence battery contains content that generally reports lower levels of differential scoring between racioethnic subgroups (i.e., measures of fluid intelligence including working memory), but it is important that data are reported on this key aspect of the test. Therefore, it is hypothesized that:

> *Hypothesis 4:* The traditional psychometric intelligence battery, which requires previously learned knowledge and entrenched problem solving techniques, will have a larger effect size compared to the neuropsychological intelligence battery, which does not rely on prior knowledge.
>
> *Hypothesis 5*: The traditional psychometric intelligence battery, which requires previously learned knowledge, will have a larger effect size compared to the alternative psychometric intelligence assessment, which does not rely on prior knowledge.

It is unknown if the alternative psychometric assessment used in the current study, a shortened version of the Siena Reasoning Test described above, will elicit subgroup mean differences, however there is no reason to expect that it will based on the prior research (e.g., Ferreter et al., 2008). The neuropsychological intelligence battery uses tests of fluid intelligence which typically have Black/White differences between *d* = .3 and *d* = .5, which is larger than what the Siena Reasoning Test previously reported. Therefore, it is hypothesized that:

> *Hypothesis 6*: The neuropsychological intelligence battery, which relies on information processing and working memory skills, will have a larger effect size compared to the alternative psychometric intelligence assessment.
>
> *Hypothesis 7*: Within the neuropsychological intelligence battery, the verbal fluency assessment, which relies on verbal ability, will have a larger effect size compared to the conditional associative learning and working memory assessments, which rely on information processing and working memory.

Now, the imperative research question is whether or not differences in scores observed on traditional psychometric intelligence tests equate to differences in brain functioning. Psychometric intelligence assessments do not have a method of evaluating cognitive functioning as it relates to performance other than mean score differences on tests of fluid and crystallized intelligence (recall Duncan and colleagues, 1995, work with the frontal and posterior patients), however neuropsychological techniques do have methods to determine cognitive functioning. The next section will examine the construct validity of the two main approaches to intelligence used in the current study and it will delineate why a brain-behavior link is important for establishing validity.

Validity Concerns for Psychometric Intelligence

Construct validity hinges on a few key questions: What is the test measuring? What additional information or constructs could the test be measuring? What is the test *not* measuring that should be included? Does variation in the standing on the construct produce variation in the response behavior and underlying psychological processes? That is, if a person scores in the 75th percentile on a test of *g*, does that person have lower cognitive functioning in the brain than a person who scores in the 90th percentile? There continues to be limited empirical evidence which demonstrates that differences in cognitive functioning in the brain result in differences in scores on psychometric assessments. Some neurological research has used persons with brain injuries and measured performance on psychometric intelligence measures, which often show differential outcomes based on the location of the injury (e.g., Duncan et al., 1995). To restate a previous theme, this type of evidence is the basis for a link connecting cognitive functioning in the brain to intelligent performance on psychometric assessments. This link is construct validity (Borsboom, Mellenbergh, & van Heerden, 2004; Embretson, 1983).

Of all the questions presented above, Borsboom and colleagues (2004) would argue that the most important questions to ask are *does the attribute exist* and *do variations in the attribute causally produce variations in the measurement outcomes* (p. 1061). These questions are very simple, but absolutely necessary for determining if an intelligence test is valid. The validation process is concerned with linking Test A to Attribute Z. Thus far, there is little evidence verifying that variations in cognitive functioning causally determine variations in the psychometric intelligence test scores. Typically, the psychometric approach to intelligence uses correlations with other psychometric intelligence assessments as evidence of validity. Positive manifold, the positive correlation between performance scores on different psychometric intelligence assessments, is commonly used as validation evidence (Spearman, 1927). Borsboom and colleagues argue that thinking of validity in terms of covariation (i.e., correlations) is flawed because of range restriction. An example similar to the one they provided is to imagine if 100 pieces of string all identical in length were measured by one ruler. The correlation between the lengths of strings and the ruler measurements would be zero (i.e., due to the fact that there is no variation in the length of the string) and the ruler would be incorrectly deemed as not valid for taking measurements. The emphasis on causality between attribute and outcome, rather than correlational relationships, is critical to their conception of validity: "The causal account says that if there are differences in the attribute, then these will produce differences in the measurement outcome. However, if there are no differences in the attribute, then no differences in the measurement outcomes are expected" (p. 1066). First, and foremost, the lack of evidence of brain-behavior connections is the main weakness of the psychometric approach to intelligence as it relates to construct validity.

Construct validity is also concerned with the operationalization of a variable. That is, how is the construct defined and what measures best capture that definition through human behavior? It was written previously that there are many conceptualizations of "intelligence", yet there is no agreement on one working definition to use in research (Neisser et al., 1996). The lack of consensus has resulted in researchers using their own understandings of intelligence and the operationalizations of the construct are as varied as the definitions. The definition problem and operationalization problem are one in the same given the detriment of this chain reaction: with many definitions of intelligence, the construct is operationalized in many different ways, thus there are various assessments which are *not* measuring the same abilities (i.e., a lack of construct validity). One of the greatest concerns is that all scores, which represent different abilities, are falling under the guise of “intelligence”. Thus, scores from a test of *g*, a working memory exercise, a quantitative ability measure, and a memory test, all can be considered measures of intelligence even though they assess drastically different aspects of cognitive functioning. Knowing that psychometric assessments measure many different areas of cognition, the content of the tests is also targeted as a limitation of psychometric intelligence’s construct validity.

A third weakness of construct validity for the psychometric approach to intelligence is that psychometric measures yield inconsistent subgroup mean differences. There are several measures that result in disparate mean scores for different racioethnic groups but some measures do not and it is difficult to know with certainty if these differences are real. There is a growing body of literature (Fagan & Holland, 2002, 2007; Freedle & Kostin, 1997) which provides empirical evidence that some of the observed differences are due to the measure that is used. Researchers have demonstrated that by modifying the intelligence measure, the subgroup differences can be minimized or eliminated (Fagan & Holland, 2002; Ferreter et al., 2008). In other words, the differences that are present are because of *how* intelligence is assessed, not differences in the cognitive functioning of different groups. Given that mean differences appear when intelligence is measured in different ways directly impacts the construct validity of the psychometric approach to intelligence. Currently, it is the stance of the APA task force that the evidence of genetic differences between White/Caucasian and Black/African American test takers fails to explain the mean score differences of these groups (Neisser et al., 1996, p. 95). Without variations in cognitive functioning, there is no reason to hypothesize that subgroup mean differences are due to the characteristics of the test takers. Rather, a major tenet of this research study is that it is the characteristics of the measure that cause variations in test scores.

The neuropsychological approach to intelligence has a considerable benefit in its favor for establishing construct validity: neuroimaging. Neuroimaging removes many of the inferences needed compared to other approaches to intelligence research. When a participant completes a task and the corresponding neuroimages display activation in the prefrontal cortex (or the right occipital lobe, for example) it is a necessary tether back to the intelligence construct. Using neuroimaging, even in conjunction with psychometric intelligence measures, can help advance construct validity of both neuropsychology and psychometrics.

Using neuropsychology to improve construct validity. Neuroimaging techniques have shown to be reliable and valid instruments in identifying areas of the brain that are activated during particular tasks as well as monitoring blood and glucose levels. However, the process of carrying out an EEG, PET, or fMRI is extremely expensive, time consuming, and cumbersome. Whether an individual has sensors secured to his scalp, or if he is required to swallow or be injected with radioactive isotopes, practicing neuroimaging is an involved process and harbors some risk.

To use an fMRI or an EEG to measure intelligence in a selection or academic setting would be impractical if not impossible. Knowing the need for a practical neuropsychological instrument, batteries of computerized neuropsychological assessments were developed. One battery in particular was based on clinical studies using non-human primates with brain lesions, human patients with neurological dysfunctions, and neuroimaging of normally functioning humans. The assessments included in the neuropsychological battery associate “as specifically as possible with dorsolateral prefrontal cortical function” (Higgins et al., 2007, p. 301) and incorporate the executive functioning tasks described above (i.e., working memory, self-ordered pointing task, conditional associative tasks). Human patients and non-human primates show causal links between conditional associative tasks and frontal lobes of the brain. Past research demonstrated that human patients who had portions of their right or left frontal lobe surgically removed for the treatment of epilepsy experienced differing cognitive deficits (Petrides, 1985, 1990). Patients who had right frontal lobe excisions were subject to nonverbal impairments, and patients with left frontal lobe excisions experienced verbal impairments. Additionally, using PET neuroimaging on normally functioning adults confirmed activation in the left frontal cortex during nonspatial (i.e., verbal) conditional associative tasks (Petrides, Alivisatos, Evans, & Meyer, 1993). Compared to normal control subjects, patients who experienced left and right frontal lobe excisions performed poorly on this the self-ordered pointing task (Petrides & Milner, 1982). The other working memory tasks include recency tasks in which words are flashed and the participant needs to recall which word appeared more recently. Prior review of the literature has empirically linked working memory with activation of the frontal lobes of the brain (specifically the dorsolateral prefrontal cortex or DLPFC; Conway et al., 2003). Additionally, damage to the left frontal lobes has demonstrated performance deficits on verbal fluency tasks (Milner & Petrides, 1984). In further research, functional imaging (e.g., PET) shows activation in the left DLPFC during verbal fluency tasks (Frith, Friston, Liddle, & Frackowaik, 1991) for healthy normal participants.

Neuropsychological assessments have existed for many decades, however administration was a very manual and time consuming process. Stimuli were often presented on large cards and the individual's responses were recorded by the administrator or an assistant. All metrics were subject to human error, especially those related to minute measurements such as reaction time. Now, neuropsychological assessments are computerized which enables faster and automated administration and more accurate data collection.

To be clear, this study does not use neuroimaging techniques. The tasks used as part of the neuropsychological battery have already been empirically linked to the dorsolateral prefrontal cortex. Petrides and her colleagues (Milner & Petrides, 1984) first used paper/pencil versions of the tasks with frontal-lobe lesion patients to document performance deficits. Later they used neuroimaging (e.g., PET and MRI) on human participants to complete the self-ordered pointing task and conditional tasks, which showed increased activation in the dorsolateral frontal cortex (Petrides et al., 1993). It should be noted that this study assumes equivalence between Petrides' work using paper stimuli and Higgins and colleagues' (2007) work using computerized stimuli. This inference is sound given the reviews on paper/pencil and computerized test equivalence. Mead and Drasgow's (1993) meta-analysis found no medium effects for tests similar to those used in the present study (i.e., tests similar the GRE). More recent reviews (Noyes & Garland, 2008) reiterated Mead and Drasgow's findings and added that medium effects were also not found on personality inventories. In one study (Lukin, Dowd, Plake, & Kraft, 1985), participants took three personality inventories. No significant medium differences were reported and computer administration was preferred by 85% of the participants.

Neuropsychological batteries enable researchers to measure performance on tasks that require prefrontal cortical activity without the inconvenience, cost, and potential harm to human participants. Neuropsychological assessments are frequently used in clinical research, but their use in a selection context or in psychometric research is less common. The current research used a variety of neuropsychological assessments encompassing executive functioning, working memory, and attention which have been shown to activate the DLPFC. The focus of the current study is on neuropsychological measures that activate only the frontal lobes, not other areas of the brain responsible for intelligent behavior. The brain-behavior link is activated when completing the neuropsychological battery used in the present study. Using a neuropsychological battery may be a viable alternative in a selection context. However, there are questions about how the neuropsychological battery will relate to the psychometric intelligence battery and assessment. That is, are psychometric intelligence tests and neuropsychological intelligence assessments related (i.e., do they show construct validity)? To help answer this question, it is hypothesized that:

> *Hypothesis 8*: The neuropsychological intelligence battery will be positively and moderately correlated with traditional psychometric intelligence battery.
>
> *Hypothesis 9*: The neuropsychological intelligence battery will be positively and moderately correlated with the alternative psychometric intelligence assessment.

Additionally, because the alternative psychometric assessment employs similar cognitive functions as the neuropsychological battery (e.g., a dependence on working memory) it is plausible that these two scores will be related. Therefore it is hypothesized that:

> *Hypothesis 10*: The neuropsychological intelligence battery will have a higher intercorrelation with the alternative psychometric intelligence assessment compared to the intercorrelation with the traditional psychometric intelligence battery.

Summary

The previous sections have reviewed the current literature on psychometric and neuropsychological approaches to intelligence, racioethnic differences on general intelligence and other subdimensions of intelligence (i.e., verbal, quantitative, information processing). Measures of psychometric intelligence have common variance, *g*, which is an underlying factor of performance on a variety of tasks (Spearman, 1927). That is, one who performs well on a task has a high probability of performing well on a different task partially due to a general factor of ability. However, it is widely accepted that measures of general intelligence have large mean score differences for Black/African American and White/Caucasian test takers (*d* = 1.0; Hough et al., 2001). Furthermore, questions regarding the validity of *g* have emerged in the intelligence literature (Goldstein et al., 2009). To examine *g*, it is difficult to say that *g* is intelligence. In fact, *g* is not a mechanism in the brain or a neural process, but rather a stable statistical artifact of shared variance (Richardson, 2005). If *g* is statistical variance, where does it reside in a human? And how does *g* help or inhibit an individual perform on various cognitive ability tasks? The benefit of expanding intelligence research into the neuroscience domain is that it helps to answer the “where” and “how” of cognitive functions (Duncan, 2005).

Neuroscientists have successfully identified areas in the brain that are likely responsible for intelligence (Conway et al., 2003; Duncan et al., 1995; Kane & Engle, 2002; Prabhakaran et al., 2000). Case studies purported that crystallized intelligence is located in the posterior region on the brain abaft the central fissure (Duncan et al., 1995) and fluid intelligence is located in the frontal cortex (Kane & Engle, 2002). Functional imaging techniques provide support for prefrontal cortical activation (i.e., the dorsolateral prefrontal cortex) during working memory tasks (Prabhakaran et al., 2000), which are very similar, if not identical, to fluid intelligence (Kane & Engle, 2002). In the last 25 years, great gains were made in identifying where specialized cognitive functions reside in the brain. Certain cognitive functions have been mapped to regions of the brain other than the frontal cortex. For example, verbal comprehension is located in the left superior temporal lobe and motor control is located in the basil ganglia (Horton & Wedding, 1984), however the focus of the present study is on the functions that reside in the prefrontal cortex. Neuropsychological intelligence may be the frontier in intelligence testing however much more research is needed. Blending different approaches to intelligence and comparing each approach’s predictive ability, subgroup mean differences, and construct validity is a relatively new practice in the intelligence research domain. The current study will use neuropsychological and psychometric intelligence batteries as well as an alternative psychometric intelligence assessment to answer the research questions posed in the Introduction and to test the hypotheses offered in the Literature Review. The methodology for the current study is presented in the following chapter.

Chapter 2: Method

Participants

A total of 312 undergraduates enrolled at Baruch College participated in the study. Approximately 56% were female and the average age was 22.1 years old (median age = 20 years). The hypotheses are based on theories about performance differentials between Black/African American, Hispanic, and White/Caucasian test takers, therefore only these three racioethnic groups were eligible to participate in the study. Students who self-identified their racioethnic group as White/Caucasian were the largest group (52%), followed by Hispanics (31%) and Black/African Americans (17%). Students participated in exchange for course credit.

Neuropsychological Measures

Neuropsychological intelligence was measured using a computerized battery of seven neuropsychological assessments. The seven assessments are verbal and nonverbal and each assessment is categorized into one of three topics. The first three tasks (spatial conditional associative task, nonspatial conditional associative task, and go/no-go) are categorized as conditional associative learning tasks. Conditional tasks require the participant to learn behavioral rules (e.g., If Prompt A, then Respond with Behavior Y; if Prompt B then Respond with Behavior Z). The basic rules are the same for spatial and nonspatial conditional associative tasks but feature different stimuli (i.e., nonspatial tasks typically have a verbal element and spatial tasks frequently use shapes). The go/no-go is also a conditional task (e.g., If Prompt A, then Respond with Go; if Prompt B then Respond with No-Go). The next three tasks (self-ordered pointing task, randomization task, and recency task) are categorized as working memory tasks. Working memory is the ability to store and integrate information with current thought and was previously discussed in Literature Review. The last task is a word fluency task. Verbal fluency tasks typically require the participants to generate words based on a given rule (e.g., words that start with the letter M). Each of the seven tasks is described in greater detail below.

Spatial conditional associative task. For this task five circles and five squares are presented in a fixed spatial array. Each square is associated with a corresponding circle. For each trial one circle is highlighted and the participant must pick the corresponding square that he or she thinks is associated with the highlighted circle. The computer will indicate if the participant chose correctly either by displaying “Correct” on the screen or displaying “Wrong” and allowing the participant to select another circle. The task is completed either after 10 consecutive correct trials or 100 trials, whichever condition occurs first. The participant will complete a second spatial conditional associative task that uses the same instructions but differs in the shapes that are used and the spatial arrangement. The score for spatial conditional associative ability is a raw error score that equates to the number of trials completed across both tasks. Both tasks are untimed.

Nonspatial conditional associative task. The nonspatial conditional associative task is similar to the spatial task described above but instead of shapes it uses words in one task and nonsense words (e.g., *egtao*) in a second task. The first nonspatial conditional associative task instructs the participant to learn associations between five cue words and five target words. For each trial one cue word is presented with five target words arranged around it in a circle. The participant selects which target word is associated with the cue word. The computer will indicate if the participant chose correctly either by displaying “Correct” on the screen or displaying “Wrong” and allowing the participant to select another target word. The task is completed after 10 consecutive correct trials or 100 trials, whichever condition occurs first. The participant will complete a second nonspatial conditional task that uses nonsense words. Meaning the participant will be instructed to learn associations between five nonsense cue words (i.e., *hogna*) and five nonsense target words (e.g., *egato*). The nonsense task continues exactly the same as the first "real words" task. The score for nonspatial conditional associative ability is a raw error score that equates to the number of trials completed across both tasks. Both tasks are untimed.

Go/no-go. Participants are instructed to click a button when certain letters appear and not to click when other letters appear. One trial consists of one flashing letter on the screen which appears for 2 seconds. The computer will notify participants who make a correct response (e.g., clicking or not clicking for the letter as indicated in the instructions) by displaying "Good" on the screen. The task is completed after 20 consecutive correct trials or 200 trials, whichever condition occurs first. The error score for go/no-go decision making ability is the number of trials completed. This task is untimed.

Self-ordered pointing task. Each task begins with 12 stimuli arranged on the screen. The participant is instructed to click on each of the 12 stimuli only once. After the participant makes a selection (i.e., clicks on one stimulus object) the arrangement of the stimuli changes immediately. The participant is allowed only 12 selections for each task. The participant will complete four self-ordered pointing tasks with different stimuli (e.g., real words, nonsense words, abstract pictures, and familiar objects). The participant is finished after completing all four tasks (i.e., making 48 total selections). The score for self-ordered pointing ability is the number of unique stimuli selected across all four tasks (i.e., 48 – number of errors). All four tasks are untimed.

Randomization task. The randomization task instructs the participant to provide a random sequence of letters starting with one letter and ending four characters away. For example, the instructions could read “Please randomize the letters from J to M” which is a four letter span. If the participant successfully provides the correct answer (e.g., one possible correct answer is M, J, L, K) then the next task will have a letter span that is lengthened by one character (e.g., 5-letter span). If the participant omits a letter or provides a sequenced string of letters (e.g., K, L, M), he or she is allowed another trial on a letter span of the same length (e.g., 4-letter span). The participant must correctly complete two 4-letter span practice trials before beginning the randomization task for a score. The task ends if the participant successfully completes all trails (4-letter to 14-letter spans) or when the participant fails on any letter span twice in a row, whichever condition occurs first. Even if the participant needed two trials for each letter span, the number of trials cannot exceed 22. The randomization ability score is the maximum length of letter span successfully completed. This task is untimed.

Recency discrimination task. For each trial, the computer will present a sequence of six or eight nouns (depending on the trial), each for 800 ms. For example, the words could be “CAR, LAMP, PICTURE, BRICK, NECKLACE, TREE”. After the entire sequence is shown, two words are shown on the screen simultaneously and the participant must pick which word appeared most recently. If the two words shown were BRICK and LAMP, then the correct answer is BRICK. The task is complete after the participant finishes 22 trials (8 trials with 6 words, 14 trials with 8 words). The recency ability score is the number of trials answered correctly (i.e., number of trials correctly answered with the more recent word). This task is untimed.

Word fluency task. This task requires the participant to produce as many words that start with the letters “ST”. Participants will be instructed not to use inflected form of words (e.g., plurals of words, the past tense). Participants generate words with the point-and-click method using the computer mouse to click at on-screen letters listed in alphabetical order. The word fluency ability score is the number of valid words generated. STREET would be a valid answer but STREETS would not also be accepted. There is only one trial for the word fluency task. Participants have 5 minutes to complete this task.

Neuropsychological intelligence scoring. All seven neuropsychological assessments are computerized and the battery’s owner, ExamCorp, will provide a data feed with each participant’s average standardized neuropsychological intelligence score, as well as standardized subdimension scores for all subdimensions.

Psychometric Measures

Traditional psychometric intelligence battery. The psychometric battery mirrors typical psychometric tests (e.g., SAT, ACT) in that it includes both verbal and quantitative components. Assessments from the O*NET Ability Profiler (formerly called the General Aptitude Test Battery or GATB) measured a participant’s psychometric intelligence. The quantitative assessment includes word problems that necessitate basic mathematical computations (i.e., addition, subtraction, multiplication, division). A practice arithmetic question is “Harry spends 1/3 of his monthly income on rent. He earns $1,560 per month. How much does he pay for rent?” There are four response options (A.

$460; B. $490; C. $530; D. $560) and a fifth option is always included that reads “E. none of these.” The quantitative assessment contains 18 questions and each participant was given 20 minutes to complete the task. Each question has only one correct answer. Quantitative ability score is total number of questions answered correctly. Higher scores indicate higher quantitative ability.

The verbal assessment is a vocabulary test that asks participants to pick two words (out of four) that are “either *most nearly the same* in meaning or *most nearly the opposite* in meaning.” The response options are the same for every question: “A - B”, “A - C”, “A - D”, “B - C”, “B - D”, and “C - D”. One practice question is testing the participant’s recognition of synonyms, which reads:

A. big
B. large
C. dry
D. slow.

“Big” and “large” comprise the only synonym given all four words, therefore the participant must select the response option that reads “A - B” to answer the question correctly. The other practice question tests the relationship between antonyms. The verbal ability assessment contains 19 questions and each participant was given 8 minutes to complete the task. Each question has only one correct answer. Verbal ability score is the total number of questions answered correctly. Higher scores indicate higher verbal ability.

Alternative psychometric intelligence assessment. The Siena Reasoning Test-25 (SRT25) contains 25 analytic reasoning items including verbal analogies containing nonsense words, visiospatial analogies using figures and shapes, and general reasoning items. All items are nonentrenched because they are stylistically unique and it is unlikely that these types of items have been encountered before. For the analogy items the question stem is three pieces of information needed to solve the analogy. An example of a nonsense verbal analogy item is:

- A GATH resembles a SHET but is heavier.
- A SHET resembles a COUCH but is heavier.
- A MUNT resembles a LAMP but is heavier.

Respondents are asked if statements such as, “A GATH is heavier than a SHET” and “A COUCH is heavier than a GATH” are true or false. Visiospatial items are similar in that three statements are presented (e.g., Figure 1 is less durable than Figure 2, Figure 3 is more durable than Figure 2). Questions based on the provided visiospatial information are asked with three available response options: true, false, and cannot be determined. General reasoning items include letter sequencing, numerical reasoning, and one nonverbal item that is a 3 x 3 matrix of shapes. The matrix has an embedded pattern but one of the nine shapes is missing. The participant is to decide which of the five response options (i.e., shapes) best fits in with the matrix’s pattern. Each question has only one correct answer and the total reasoning ability score is the total number of questions answered correctly. A higher score on the SRT25 indicates greater reasoning ability. The SRT25 technical report noted that the average score for the SRT25 is 18.9 (*SD* = 3.47; Range = 4-25) and it has acceptable internal consistency (• = 0.72). There was a 25 minute time limit for the SRT25.

Test Attitude Survey. The Test Attitude Survey (TAS) created by Arvey, Strickland, Drauden, and Martin (1990) measures participant’s attitudes on a variety of dimensions including test ease, motivation, belief in tests, and general need achievement. Not all topics are pertinent to the current study therefore only the following five dimensions were included. First, the motivation dimension contains 10 items and has acceptable internal consistency (• = 0.85). An example motivation item is “I tried my best on this test.” Second, the lack of concentration dimension contains four items and has acceptable internal consistency (• = 0.78). An example lack of concentration item is “I found myself losing interest and not paying attention to the test.” Third, the belief in tests dimension contains four items and has acceptable internal consistency (• = 0.71). An example belief in tests item is “This test was a good reflection of what a person could do in the job.” Fourth, the test ease dimension has four items and Arvey and colleagues report low internal consistency (• = 0.56). An example test ease item is “I found this test interesting and challenging.” The fifth dimension is external attribution and contains 5 items. Arvey and colleagues report low internal consistency for this dimension (• = 0.58). An example external attribution item is “I became fatigued and tired during the testing.” The five sections contain 27 items of which six items are reversed scored. An example of a reversed scored motivation item is “I just didn’t care how I did on this test or tests”. All items on this measure are rated on a seven-point scale with the anchors of “Strongly disagree” and “Strongly agree”. Reversed scored items were changed prior to the analysis of the data. Each TAS was untimed.

The other dimensions of the TAS include questions that are irrelevant to the current study. These topics include comparative anxiety (e.g., “During the test, I found myself thinking of the consequences of failing.”), general need achievement (e.g., “Once I undertake a task, I usually push myself to my limits.”), future effects (e.g., “Scores from this test will probably affect my future.”), and preparation (“I prepared a lot for this test.”). These questions are irrelevant because participants cannot prepare for the study and are told repeatedly that performance during the study will have no consequences on their academic standing nor will it affect their future in any way. Comparative anxiety and general need achievement topics are not included because these constructs are not pertinent to the present study. The items for each dimension of the TAS are cumulative. Higher scores indicate more favorable or stronger attitudes within a particular dimension. The shortened 27-item TAS is located in Appendix A.

Personality. The neuropsychological battery contains a Big-5 personality assessment called the Five-Dimensional Temperament Inventory (FDTI) adapted from Goldberg’s (1992) Trait Descriptive Adjectives. Rather than a 5-point Likert Scale, each item is a bipolar visual analogue scale. At each pole, there are three adjectives. The words at one pole are drawn from a list of adjectives that load on the positive end of the underlying factor, whereas the adjectives at the opposite pole load on the negative end of the factor. For example, a typical extraversion item might have *Unadventurous, Shy, Withdrawn* on the left end of the scale and *Active, Daring, Talkative* on the right end of the scale. Above the scale sits the question “What point on the scale best describes you?” To enter his or her response, the participant clicks on the scale using the mouse. The personality assessment was untimed.

Demographic characteristics. Participants were asked to self-identify their racioethnic group, sex, age, country of birth, socioeconomic status, primary spoken language, GPA, and SAT math and verbal scores. The demographic questionnaire was untimed. The demographic questionnaire is located in Appendix B.

Procedure

Sessions were conducted in private computer labs within the Baruch Campus. When the participants arrived, they were seated and given the consent form. The experimenter first presented a general overview of the procedure. The neuropsychological and psychometric intelligence batteries were randomly counterbalanced, meaning that half of all sessions started with participants taking the neuropsychological intelligence battery, and the other sessions started with the psychometric assessment. The purpose of switching the order of administration is to ensure the results are not due to administration order. The shortened version of the Test Attitudes Survey was administered twice: once after the neuropsychological intelligence battery and once after the psychometric intelligence battery. Next, the SRT25 and demographic questionnaire were completed. All participants began the neuropsychological intelligence battery at the same time and progress through the seven assessments and personality assessment at their own pace and then were instructed to complete the TAS. All participants started the psychometric battery, its corresponding TAS, and the SRT at the same time. The experimenter enforced time limits (when appropriate).

Planned Analyses

Correlations and regression analyses. To test Hypotheses 1a to 2c, which are all about different predictors positively predicting academic performance, the ordinary least squares regression technique was used. Pearson’s correlations were calculated to determine the relationships between all dimension variables in the study. As a reminder, the predictors for the regression analyses are the neuropsychological intelligence battery, the traditional psychometric intelligence battery, an alternative psychometric intelligence assessment (SRT25), and the conscientiousness personality dimension. The outcome variable is GPA.

Hypotheses 3a and 3b and 10, concerning the magnitude of the predictive validity coefficients will rely on the aforementioned Pearson correlations. A test for dependent correlations (see Cohen, Cohen, West, & Aiken, 2003) was used to determine if neuropsychological intelligence has a statistically significant larger predictive validity coefficient compared to other predictors (Hypotheses 3a and 3b). Hypothesis 10 will test the difference between intercorrelations of the neuropsychological intelligence assessment. The statistical significance of the full regression equation and the coefficients as well as the variance explained by the entire model will also be examined. Incremental variance explained was reported as a comparison to the results for the relative importance analysis.

Hypotheses 4 through 7 are all concerned with effect sizes on the various predictor variables (i.e., the intelligence assessments). Effect sizes can be quantified using Cohen's (1988) *d* statistic in order to put the differences onto a common metric. To calculate *d*, the comparison group's mean is subtracted from the majority group's mean and divided by the pooled standard deviation (e.g., a weighted average of the two groups' standard deviations). Currently there is no test to determine if two effect sizes are significantly different. Cohen (1988) reported that an effect size of .2 to .3 was small, .5 was medium, and .8 was large. For the purposes of the current study, the hypothesis was supported if the magnitude between effect sizes for different racioethnic groups is .30 or larger.

Hypotheses 8 and 9 predict moderate and positive correlations (i.e., $p < .05$) between the neuropsychological battery and the a) psychometric intelligence battery and

b) alternative psychometric intelligence assessment. As mentioned above, Pearson's correlations were calculated and reported for all predictor and outcome variables.

Relative Importance Analysis. There is one research question posed in the introduction that has not yet been address. The current study plans to answer the question, how relatively important are measures of neuropsychological and psychometric intelligence and personality for predicting performance? Although it falls under the domain of questions pertaining to predicting performance, it was offered without corresponding hypotheses. No research has explored the relative weights of these types of predictors, therefore the analysis will be strictly exploratory in nature.

Regression analyses were used to test other hypotheses, however there are two limitations of using multiple regression as the primary statistical technique to make interpretations about the relative importance of the tests for predicting performance outcomes. The first limitation is regarding one assumption of performing multiple regression which is that two or more predictors cannot be highly correlated. The proposed psychometric intelligence battery has never been compared to the neuropsychological battery. However, previous research has demonstrated that measures of neuropsychological intelligence are correlated with psychometric measures of intelligence and SAT scores, both which are common predictors of academic performance (Higgins et al., 2007). The strength of the correlations varied between studies (Study 1 with Harvard University females: $r = .20$, $p < 05$; Study 2 with University of Toronto students $r = .45$, $p < .01$). Hence, it is important to find a statistical technique that takes into account correlated predictors. Another limitation of regression is the utilization of the output. If a predictor has a small beta coefficient it cannot be assumed that the variable is not a “good” predictor. The magnitude of beta coefficients (and the incremental R^2 coefficients) largely depends on the order in which the predictor variables are entered into the regression equation. Therefore, an alternative statistical analysis, called relative importance analysis, will be conducted in conjunction with multiple regression to determine the relative importance of the predictor variables included in the present study. An elaboration of the definition and various techniques used in relative importance analysis follows below.

Relative importance is defined as the “proportionate contribution each predictor makes to R^2, considering both its unique contribution and its contribution when combined with other variables…” (Johnson, 2001, p. 232). A recent review of relative importance techniques considered several methodologies to obtain relative weights and importance of predictor variables (Johnson, 2004). A sample of the relative weights techniques include dominance analysis (Budescu, 1993), the product of the correlation and its standardized regression coefficients (Hoffman, 1962), and orthogonal transformations (Gibson, 1962). Each relative weights technique had considerable limitations that did not take into account the problem of correlated predictors or the arduous calculations necessary to complete the analysis. Johnson (2001) details some of the limitations for these other relative importance techniques. For example, dominance analysis determines if predictor variables can be rank ordered. This analysis requires the computation of incremental R-squares for all possible submodels. This method is too cumbersome and time-consuming to execute if there are more than five or six predictors (Linderman, Merenda, & Gold, 1980). In another example of limitations, consider the product of the correlation method which multiplies the zero-order correlation and its standardized regression coefficient

(Hoffman, 1962). This method may not be feasible because the product can either be zero or negative even if the variable is a strong predictor of the outcome. Johnson’s (2000) relative weights analysis (RWA) takes into account the problem of correlated variables and it relatively simple and quick to calculate.

Many organizations want to know what predicts a particular organizational variable (i.e., employee satisfaction). Typically there is a set of predictors of a particular outcome and these predictors may or not be related (i.e., correlated) to each other. Predictors may include employee engagement, perceived organizational justice, benefits, recognition, inclusion, diversity, and supervisor effectiveness. At times it is desirable to know how much each predictor contributes to explaining the outcome. Although all of the variables listed above predict employee satisfaction to some degree we know that employee engagement and supervisor effectiveness typically are the strongest predictors of satisfaction (i.e., they account for the most variance). Johnson’s RWA technique is useful for determining what percentage of variance is explained by each predictor variable, even when predictors are correlated (e.g., we know that employee engagement and supervisor effectiveness are highly correlated). RWA is appropriate for the current analysis to determine the relative importance of neuropsychological intelligence, psychometric intelligence, and personality variables in predicting outcomes of scholastic aptitude (as measured by GPA).

One benefit of using RWA with multiple regression is that it provides rank orderings of predictor variables on their relative importance in contributing to the percent of variance explained (R^2). Although a predictor may have a smaller beta weight as a result of regression, RWA provides a more accurate interpretation of that variable’s importance in predicting the outcome in relation to other variables included in the model. Another benefit of using RWA is that it has been found to be more reliable than multiple regression, meaning the error is less for relative weights than for traditional beta coefficients. Commenting on a Monte Carlo study in which least-squares regression was compared with RWA within samples from a population (see Oswald, Johnson, and Oliver, 2000), Johnson (2001) writes that “relative weights were consistently less biased than were least-squares regression coefficients, in that the mean sample-based values were closer to the population values” (p. 245). He also notes that the relative weights were less variable than least-squares regression across the simulation. Therefore, least-squares multiple regression was included as a basis for comparison with past research (Higgins, et al, 2007), however, RWA is also included because 1) relative weights have never been assigned to neuropsychological intelligence measures in comparison to psychometric intelligence measures when predicting academic and cognitive ability performance therefore providing a unique contribution to the literature, 2) there is a high likelihood that the neuropsychological and psychometric intelligence batteries will be correlated, which would violate an assumption of using least squares multiple regression, and 3) when comparing Johnson’s RWA to other relative weights techniques (e.g., Budescu’s dominance analysis) using the same data, RWA produced nearly identical results.

Johnson’s (2000) relative weights analysis is performed in three steps. Johnson notes that “relative weights are calculated by first obtaining the best-fitting (in the least squares sense) set of orthogonal variables” (p. 235). Orthogonal variables can be calculated directly from a correlation matrix. The correlation matrix contains the outcome or criterion variable in the first column and any number of predictors in the remaining columns. Next, transform (i.e., regress) the original (correlated) correlation matrix onto the orthogonal (uncorrelated) matrix. Then regression coefficients are assigned to the orthogonal predictors (not the original variables). Lastly, the relative weights, called epsilons (•), are the outcome of the squared regression coefficients from the two sets of variables. RWA has never been performed on these types of data, therefore the relative importance analysis will be exploratory in nature.

Chapter 3: Results

Missing data. The sample sizes for different comparisons performed in the analysis vary due to unforeseen circumstances during data collection. Sporadic computer network firewall issues occurred in two out of three computer labs where the study was conducted. This issue prevented 65 cases of computerized data from being recorded, therefore the sample of neuropsychological battery and personality data is reduced (n = 247) compared to the overall sample (N = 312). In one session, the SRT was unable to be administered due to time constraints, therefore a total of 306 completed SRT records are available. Finally, all questions in the study were optional and some students choose not to report their GPA (n = 292) which reduced the number of complete records at the item level. No particular group withheld demographic information more than another and samples with missing data are representative of the total sample. Sample sizes are noted in the tables.

Descriptive statistics. Table 1 contains the descriptive statistics for the predictor and criterion measures used in the study. Reliability estimates were not available on the neuropsychological battery and personality factors because only dimension level scores were provided by the test vendor. As can be seen in the table, available estimates of internal consistency were within an acceptable range. The scores on the traditional psychometric intelligence battery demonstrated acceptable internal consistency (• = 0.69) as did the nontraditional psychometric intelligence assessment (• = 0.74). A test of order-effects indicated that there was no difference whether participants received the 30-minute traditional psychometric intelligence battery first or the 90-minute neuropsychological intelligence battery first. There were no significant or meaningful differences in the outcome scores for either intelligence battery or the alternative psychometric intelligence assessment, which was taken in the last 30 minutes of the entire study (see Table 1).

The correlates of predictor, criterion, and all personality dimensions, as well as the sample sizes for each comparison, are located in Table 2. The neuropsychological battery was the only measure that was statistically significantly correlated with college GPA. All cognitive measures were significantly correlated with each other. All personality factors were significantly correlated with each other, but not with the cognitive variables.

The descriptive analysis revealed that some data violated assumptions of normality, specifically that the data were not normally distributed. The traditional psychometric intelligence battery and the alternative psychometric intelligence assessment both failed tests of normality (see Table 3). A possible floor effect was experienced on the traditional psychometric intelligence battery. Out of a maximum 18.5 points (averaged from math and verbal total scores), the mean was 12.74 (*SD* = 2.35). Six negative outliers brought the average score down and negatively skewed the data (skew = -0.42). Interestingly, there was very little variation in the mean scores of the racioethnic groups. Black/African Americans (M_B = 12.31), Hispanics (M_H = 12.81) and White/Caucasians (M_W = 12.83) all had very similar scores which is unusual for a traditional psychometric intelligence battery. Typically subgroups can vary between *d* = .5 to *d* = 1.0 (Hough et al., 2001). The alternative psychometric intelligence assessment (SRT25) has a possible total score of 25 and the overall mean was 16.67 (*SD* = 3.98). These data were also negatively skewed (skew = -0.52). There was more variation in the subgroup’s scores, however all three group’s scores were still within approximately 1.5 points (i.e., effect sizes ranged from d = -.04 to d = .35).

The mean and standard deviation of the neuropsychological battery, which are standardized scores, were lower compared to its normative sample. A normally distributed standardized score has a mean of zero and a standard deviation of 1.0. The mean for the current study’s sample was -0.34 (SD = 0.44), that is, the mean was one-third of a standard deviation below the normative group’s mean. Lastly, the dependent variable, grade point average (GPA) was also negatively skewed (skew = -1.27) and leptokurtic (kurtosis = 3.02). The majority of the participants (i.e., 92%) reported their GPA in only three response categories out of seven categories provided. The lack of variance in the dependent variable presented some problems in finding predictive relationships in the present study. A correction for indirect range restriction was considered but it did not meet the assumption of homoscedasticity (i.e., the predictor and criterion variables have the same variance at different levels of the data).

Hypothesis testing

Table 4 contains the results of Hypotheses 1a to 2c which were tested using an ordinary least-squares regression model. The criterion variable for all tests was college GPA and each hypothesis’s respective cognitive or personality factor was entered as the model’s predictor. Hypothesis 1a, that the neuropsychological intelligence battery will positively predict academic performance, was supported. It was the only variable that predicted academic performance with statistical significance ($p < .05$). Hypotheses 1b and 1c, that the traditional psychometric intelligence battery and the alternative psychometric intelligence assessment would, respectively, predict academic performance, were not supported. The results indicated that the traditional psychometric intelligence battery trended in the hypothesized direction but did not achieve traditional levels of statistical significance ($p = 0.057$). Hypotheses 2a, 2b, and 2c, that conscientiousness would predict academic performance and provide incremental validity to the intelligence batteries, were not supported. Conscientiousness was not found to be an independent predictor of GPA and it was unable to provide incremental validity in any model. Hypotheses 3a and 3b made predictions about the size of the validity coefficients from hypotheses 1a, 1b, and 1c. Hypothesis 3a, that the neuropsychological intelligence battery will have a larger predicative validity coefficient for academic performance compared to the traditional psychometric intelligence battery, was supported. A test of the difference between dependent correlations indicated that the correlation between neuropsychological intelligence and college GPA ($r = .14$) was significantly greater than the correlation between traditional psychometric intelligence and GPA ($r = .11$, $t\,(245) = 2.21$, $p < .05$). Additionally, Hypothesis 3b, that the neuropsychological intelligence battery will have a larger predicative validity coefficient for academic performance compared to the alternative psychometric intelligence assessment, was supported. The correlation between neuropsychological intelligence and GPA ($r = .14$) was significantly greater than the correlation between alternative psychometric intelligence and GPA ($r = .09$, $t\,(245) = 2.26$, $p < .05$).

Hypotheses 4 through 7 are concerned with the effect sizes of the different cognitive ability measures between demographic (e.g., self-identified gender and racioethnic) groups. Table 5 presents the mean scores by gender and racioethnic group, as well as effect sizes of the comparisons. Recall that the hypotheses were supported if the difference in effect sizes between two tests was greater than .30.

Male and female comparisons. The effect size comparisons offered in hypotheses 4 through 7 were offered mostly for the racioethnic subgroup comparisons, however, given that neuropsychological research has never reported potential score differences by sex, it is important to provide documentation of these differences. Overall, males outperformed females on all cognitive assessments except for two subdimensions of the neuropsychological battery, therefore all effect sizes favor males unless otherwise noted. Hypothesis 4, that the traditional psychometric intelligence battery will have a larger effect size compared to the neuropsychological intelligence battery, was not supported. Although the traditional psychometric battery had a larger effect size (d = .34) compared to the neuropsychological intelligence battery (d = .25), the difference was not greater than .30. However, it is important to note that the effect size found here for the traditional test battery is much lower what is typically reported in the literature (i.e., d = 1.0). Hypothesis 5, that the traditional psychometric intelligence battery will have a larger effect size compared to the alternative psychometric intelligence assessment, was not supported. The results indicated that the psychometric battery had a larger effect size (d = .34) compared to the alternative psychometric intelligence assessment (d = .09), but the difference was .25 points, therefore the hypothesis was not supported. Hypothesis 6, that the neuropsychological battery will have a larger effect size compared to the alternative psychometric intelligence assessment, was not supported. The difference between the neuropsychological battery effect size (d = .25) and the alternative psychometric intelligence subdimension effect size (d = .09) was .16. The last hypothesis concerning male and female mean differences pertained to the subdimensions of the neuropsychological intelligence battery. Hypothesis 7, that the verbal fluency subdimension will have a larger effect size compared to the conditional associative learning and working memory subdimensions, was not supported. The conditional associative learning (*d* = .36) had the largest effect size followed by word fluency (*d* = .15), and then working memory (*d* = .12).

Black/African American and White/Caucasian comparisons. The results of the Black/White test taker comparisons are a focal area of the current study. White/Caucasian test takers outperformed the Black/African American participants on most of the cognitive ability measures therefore effect sizes favor White/Caucasians unless otherwise noted. Hypothesis 4, that the traditional psychometric intelligence battery will have a larger effect size compared to the neuropsychological intelligence battery, was not supported. The traditional psychometric battery had a smaller effect size (*d* = .21) compared to the neuropsychological intelligence battery (*d* = .33). Hypothesis 5, that the traditional psychometric intelligence battery will have a larger effect size compared to the alternative psychometric intelligence assessment was not supported. The results of the study indicated that the psychometric battery had smaller effect size (*d* = .21) compared to the alternative psychometric intelligence assessment (*d* = .35). Hypothesis 6, that the neuropsychological battery will have a larger effect size compared to the alternative psychometric intelligence battery, was not supported. The neuropsychological battery had a slightly smaller effect size (*d* = .33) compared to the alternative psychometric assessment (*d* = .35). Lastly, hypothesis 7, that the verbal fluency assessment will have a larger effect size compared to the conditional associative learning and working memory assessments, was not supported. The effect size of word fluency was the smallest (*d* = -

.01) followed by working memory (*d* = .21, favoring Whites/Caucasians) and conditional associative learning (*d* = .31, favoring Whites/Caucasians).

Hispanic and White/Caucasian comparisons. In general, the effect sizes resulting from Hispanic/White comparisons were smaller compared to effect sizes from Black/White comparisons. In some cases Hispanic test takers outperformed White/Caucasian test takers (e.g., alternative psychometric intelligence assessment, word fluency task). Hypothesis 4, that the traditional psychometric intelligence battery will have a larger effect size compared to the neuropsychological intelligence battery, was not supported. The traditional psychometric battery had essentially no group differences (*d* = .01) and the neuropsychological intelligence battery differences were also small (*d* = .08, in favor of Whites/Caucasians). Hypothesis 5, that the traditional psychometric intelligence battery will have a larger effect size compared to the alternative psychometric intelligence assessment was not supported. The psychometric battery had a minimal effect size (*d* = .01) and the alternative psychometric intelligence assessment slightly favored the Hispanic test takers (*d* = -.04). Hypothesis 6, that the neuropsychological battery will have a larger effect size compared to the alternative psychometric intelligence battery, was not supported. Although the neuropsychological battery had a slightly larger effect size (*d* = .08, in favor of White/Caucasians) compared to the alternative psychometric assessment (*d* = -.04, in favor of Hispanics), in practice both effect sizes were quite small compared to what has been documented in previous research (Hough et al., 2001). Hypothesis 7, that the verbal fluency assessment will have a larger effect size compared to the conditional associative learning and working memory assessments, was not supported. The effect size of word fluency was the smallest (*d* = -

.03, in favor of Hispanics) followed by conditional associative learning (*d* = .08, favoring Whites/Caucasians) and then working memory (*d* = .10, favoring Whites/Caucasians).

The last three hypotheses are concerned with magnitude of relationships between the cognitive ability measures for the entire sample. Hypothesis 8, that the neuropsychological intelligence battery will be positively and moderately correlated with traditional psychometric intelligence battery, was supported ($r = .489$, $p < .01$). Hypothesis 9, that neuropsychological intelligence battery will be positively and moderately correlated with the alternative psychometric intelligence assessment was supported ($r = .509$, $p < .01$). Hypothesis 10, that the neuropsychological battery will have a higher intercorrelation with the alternative psychometric intelligence assessment compared to the intercorrelation with the traditional psychometric intelligence battery, was not supported. The test for the difference between two dependent correlations was not statistically significant. The correlation between the neuropsychological intelligence battery and alternative intelligence assessment ($r = .51$) was not significantly larger than the correlation between neuropsychological battery and the traditional psychometric intelligence battery ($r = .49$, t (245) = 0.37, $p > .05$).

Additional analyses. Some of the results were unexpected because they did not replicate previous findings well established in the literature. Typically traditional and alternative psychometric tests predict academic performance (Hough et al., 2001; Yusko et al., 2010) however these predictive relationships were not statistically significant in the present study. One potential cause for the inconsistent results may be due to the demographic composition of the sample, specifically, among those participants who identified as “White/Caucasian”. A closer examination of the demographics of those participants who self-identified as White/Caucasian (n=162) revealed that only 45% of respondents were born in the United States, 48% marked that English was their first language, and 36% confirmed that English was the primary language spoken in their households while growing up. In other words, the White/Caucasian sample drawn from Baruch College is probably different compared to White/Caucasian students who participate in research at other American universities, which may have lead to the atypical findings. A post-hoc comparison of only White/Caucasians' intelligence scores using the three aforementioned demographic variables as comparison groups reveal no mean differences on the neuropsychological intelligence battery nor on the alternative psychometric intelligence assessment. That is, there are no mean differences depending on if the participant was born in the US, used English as a first language, or if English was the primary language spoken in the household for those two measures. However, there were statistically significant differences on mean scores for the traditional psychometric intelligence battery. White/Caucasian participants who were born in the U.S. scored significant higher on the traditional psychometric intelligence battery (*d* = .68), as did those whose first language used was English (*d* = .73), and participants whose primary household language was English (*d* = .51; see Table 6). Although many of the respondents who were born in the United States also used English as their first language, the categories are not mutually exclusive. Of the participants born in the U.S., 16% first spoke a language other than English and 35% reported that a language other than English was the primary language spoken in their household.

Another possible reason for the inconsistent results was the shape of the distributions of the cognitive ability assessments and GPA. Several steps were taken to see if this issue could be corrected. First a log linear transformation of the criterion variable (i.e., college GPA) was performed which attempted to elicit a more normal distribution from the variable. The same hypotheses were tested on the data replacing the criterion with the transformed variable. The results were actually worse in that the neuropsychological battery no longer predicted academic performance and all other predictive models remained non-statistically significant.

The criterion variable thus far has been treated as an interval variable however, it can be argued that academic performance is ordinal data because the categories are even increments and a higher score indicates higher academic performance. Through post-hoc exploratory analysis the associations between the cognitive tests and the academic performance data were also assessed using eta (η), which is a measure of the relationship between ordinal and continuous data. In this study, eta was squared to determine the percent of variance in academic performance that could be explained by the cognitive ability measures. The results are summarized in Table 7. The results appear to be spurious due to the impractical amount of variance in GPA explained by the neuropsychological battery (e.g., 95%). These high etas are most likely due to the range restriction in the criterion variable and floor effects in the cognitive ability assessments. Given that neither strategy to overcome the limitations of the data was successful, the original analyses are used to draw conclusions about the results of this study.

Chapter 4: Discussion

The purpose of the present study was to test the predictive relationships of neuropsychological and psychometric intelligence batteries as well as an alternative psychometric intelligence assessment and a personality measure in relationship to academic performance. The study also reported subgroup mean differences and the construct validity between psychometric and neuropsychological intelligence measures. The discussion section will be presented in a similar order because it answers the research questions presented in the Introduction. Implications and limitations of the study are noted as are suggestions for future research.

Summary of Findings

Predicting performance. The first question posed in the current study is whether or not the neuropsychological intelligence battery would continue to predict performance with a heterogeneous sample. The results indicate that a neuropsychological battery is predictive of GPA in a racioethnically diverse college. These results, especially when paired with previous research which found the neuropsychological battery was predictive of academic and job performance (Higgins et al., 2007), give value to neuropsychological measures as alternate predictors compared to more traditional psychometric assessments. For decades the field of psychology has relied on a single-answer solution to predicting performance: tests of *g*. However more evidence is emerging that neuropsychological measures can predict performance, and at times they are more predictive than psychometric intelligence measures. The neuropsychological mechanisms activated in the prefrontal cortex which use working memory and attention are tied to cognitive processes needed to succeed in academic and organizational settings. It could be argued that information processing is a more critical ability than information recall (often measured in tests of *g*) in achieving high performance. Given that new and different data are becoming available it is imperative to continue to explore various predictive models of intelligence and not to rely on static solutions.

The second research question asked whether or not the other cognitive and non-cognitive measures (i.e., psychometric intelligence battery, alternative psychometric intelligence assessment, conscientiousness) were able to positively predict academic performance. Much of the literature finds a predictive relationship between psychometric intelligence and academic performance (Roth et al., 2001; Schmidt & Hunter, 1998), therefore it was unexpected that the current study did not result in similar findings. These non-significant relationships are most likely due to the data distributions on psychometric measures (e.g., they failed tests of normality). The sample may also be either low on ability or low on motivation, or both, which may have resulted in restriction of range on the psychometric measures. Unfortunately, a correction for range restriction was not possible because the data did not meet all of the assumptions for a correction. Both psychometric intelligence measures had low average scores. The traditional psychometric intelligence battery had 18.5 total possible points and the participants averaged 12.7 points. In other words, if the sample were to get a letter grade on the psychometric battery it would be a D+ (i.e., the average score was 69% correct).

In previous research, the alternative psychometric assessment positively predicted both academic and job performance (Yusko et al., 2010), however the relationship was not supported in the current study, again probably a result of the low mean score. Out of a possible 25 points, the participants averaged 16.8 points, which is 67% correct (i.e., another D+). This sample's mean score was two points lower than the normative sample (i.e., normative sample *M* = 18.85, *SD* = 3.47, unpublished SRT25 technical manual). It is important to emphasize that this study used cognitive ability assessments that could be used in an organizational setting. These assessments are designed to make distinctions among a wide array of job applicants. In a pre-employment selection context, applicants try their hardest and put forth their own maximal effort in order to perform well. A laboratory study recruits an entirely different group of respondents. Given that assessments were used that should elicit maximal performance condition but the study participants were minimally motivated to perform well may have contributed to the low scores, floor effects, and atypical findings. Additionally, it is difficult to speculate as to why the current study's sample had negatively skewed results for the psychometric measures and why the neuropsychological intelligence measures were only slightly positively skewed and considered normal. In a model of intelligence, it may be helpful to measure the participant's level of motivation and its relationship to performance.

Lastly, there was no predictive relationship between the personality factor conscientiousness and GPA, nor did conscientious add predictive validity which is contrary to previous research and meta-analytic findings (Hurtz & Donovan, 2000; Schmidt & Hunter, 1998). In general, personality factors are not able to predict performance as well as psychometric intelligence measures but a predictive relationship was expected between conscientiousness and performance. Again, it is difficult to know why the current sample did not produce a predictive relationship. One possible explanation is with conscientiousness having mixed results on the tests of normality (i.e., the Kolmogorov-Smirnov test indicated the data were borderline normal and the Shapiro-

Wilk test indicated that the data were not normally distributed), the non-normal distribution of the personality data may have contributed to non-significant predictive relationship with GPA. Another possibility is that it may be unusual for conscientiousness to be predictive of academic performance when psychometric intelligence could not predict GPA because the conscientiousness-performance relationship (e.g., $r = .20$) is a weaker than the psychometric measure-performance relationship (e.g., $r = .51$). The lack of predictive relationships in the current study was disappointing, however steps were taken to find relationships within the data but these attempts were unsuccessful. The data should be interpreted with the notes of caution listed in the sections.

The predictive relationship between the neuropsychological intelligence battery and academic performance compared to the other intelligence predictors was an important outcome. Another question posed in the current study is whether or not the neuropsychological battery was the best predictor of all the intelligence predictors. Obviously, the only significant predictor was the strongest predictor but the magnitude of the difference of prediction may be meaningful. For years, the best predictor of academic performance was thought to be a psychometric intelligence test (i.e., test of *g*). This study demonstrated that the neuropsychological battery may be a better predictor than other psychometric intelligence or personality measures. Of course, this projection is based on the few studies that have used neuropsychological intelligence measures in conjunction with psychometric measures. The results must be weighed with the amount of evidence available. However, one hope is that as more research using both approaches to intelligence becomes available, the research community will be open to the findings and consider new approaches to measuring intelligence.

The planned analysis section stated that a relative weights analysis would be performed. Relative weights are useful for determining the weight or importance of multiple predictors and how they relate to the criterion. Past research has demonstrated the neuropsychological intelligence battery, a psychometric intelligence battery, and personality factors were all significant predictors of GPA (Higgins et al., 2007). In the case of many predictors it is beneficial to know, for example, how important neuropsychological intelligence is compared to conscientiousness. However, the relative weights analysis could not be performed in the current study given that the neuropsychological battery was the only statistically significant predictor of academic performance. Relative weights analysis cannot be performed when the predictors are not significant in a regression analysis. The predictive relationship of the neuropsychological battery to academic performance holds promise, however, it cannot be considered for implementation without an examination of possible subgroup mean differences.

Subgroup mean differences. There is no known literature reporting potential subgroup mean differences on the neuropsychological intelligence battery. Examining these differences is an important contribution to the intelligence literature. Previous research samples may not have been diverse enough to look at group differences. The current study had a racioethnically diverse sample which enabled an examination of group differences. If subgroup mean differences were found in the comparisons then they must be considered before recommending the battery for use. The magnitude of the effect size should be considered as well as possible implications for a selection context (i.e., adverse impact).

The primary question pertaining to this section is whether or not the neuropsychological intelligence battery exhibited subgroup mean differences. In short, the answer is yes. That is, subgroup mean differences were found for male and female (*d* = .25), Black and White (*d* = .33) and Hispanic and White (*d* = .08) comparisons. The sex differences were unexpected but small effect sizes (*d* = .1 to .2, favoring either sex depending on if the test is verbal or quantitative in nature) were also found in previous research. Upon further examination, Hough and colleagues report an effect size of .4 (favoring males) for spatial ability measures which provides an explanation as to why the conditional associative learning tasks (spatial and nonspatial tasks included) had an effect size of nearly the same magnitude (*d* = .36). Additionally, the study's working memory effect size (*d* = .12) was larger than the meta-analytic literature for males and females (*d* = 0; Hough et al., 2001). Not all subdimensions had moderate effect sizes. For example, the random letter span tests and recency judgment tasks both report practically no subgroup differences (*d* = -.02, slightly favoring females).

Another surprising finding was the reduced subgroup differences on the psychometric intelligence measures. An effect size between 0.7 and 1.0 was expected on the psychometric measures for Black and White test takers and an effect size of 0.5 was expected for Hispanic and White test takers. The psychometric intelligence battery produced effect sizes about one-quarter of the expected size (*d* = .21) in the Black/White comparison and no group differences were found when comparing Hispanic and White test takers. All of these differences are smaller than what was found on the neuropsychological battery. These findings may indicate that the neuropsychological measures elicit larger group differences than psychometric measures. Or the results may be specific to the particular sample drawn from Baruch College. Without more research using a variety of samples it is difficult to determine the magnitude of effect sizes that can be expected from neuropsychological measures. However, from a theoretical basis, the neuropsychological measures should elicit smaller effect sizes than measures of general cognitive ability because neuropsychological measures draw upon the same or similar mechanisms that result in smaller effect sizes than general ability measures (e.g., spatial ability measures typically result in $d = .7$; memory measures typically results in $d = .5$; Hough et al., 2001).

Even though the neuropsychological effect sizes were larger than the psychometric effect sizes, the racioethnic effect sizes found on both batteries were a fraction of the size compared to typical differences in the literature. Black/African American and White/Caucasian effect sizes for psychometric assessments are typically $d = 1.0$ and Hispanic and White effect sizes are typically $d = .5$. The neuropsychological battery including all of its subdimensions had no effect size larger than $d = .35$. This particular sample also produced smaller than usual differences for the traditional psychometric measure and approximately average differences for the alternate psychometric measure. Mean score differences of this magnitude are not ideal but considering that they are much smaller than what is typically found in the literature, these measures deserve further consideration. The neuropsychological intelligence battery especially warrants further investigation because it was a significant predictor of GPA.

All researchers must be cognizant of the diversity-validity dilemma. The field is still in search of a measure (or a battery of measures) that is a strong predictor of performance while not functioning differentially for sex or racioethnic groups. There have been calls to expand the definition and operationalization of intelligence to include the contribution of several disciplines including psychometric, neurobiological, and sociological (Deary, 2001; Sternberg, 2001), however all of the work must be done while being watchful of subgroup mean differences. One way to do this is to continue to probe to find predictive assessments or battery subdimensions with reduced subgroup mean difference. For example, the neuropsychological battery had several subdimensions with varying effect sizes. One potential solution is to select a number of subdimensions with smaller effect sizes and determine if the shorter battery is still predictive. In the current study the word fluency task (d = -.01), random object span test (d = -.05, favoring Black test takers), and the conditional spatial association test (d = .13, favoring White test takers) had the lowest effect sizes (see Table 5). In a post-hoc regression analysis, these three subdimensions were tested as the new neuropsychological predictor. Unfortunately, this particular set of subdimensions was not a significant predictor of GPA. It is possible other combinations may reveal more favorable results. Finding the balance between prediction and group differences has a long and challenging history in our field, but further research is needed to find the best possible combination of factors.

Turning to another assessment of interest in the current study is the alternative psychometric intelligence assessment, or the Siena Reasoning Test-25 (SRT25). In a previous study, a different form of the SRT25 had equal predictive validity power compared to traditional psychometric assessments but exhibited very small group differences (Ferreter et al., 2008). In the current study, the SRT25 exhibited small male/female differences (*d* = .09, favoring males) and Hispanic/White differences (*d* = -.04, favoring Hispanics), but elicited a higher than expected effect size for the Black/White comparison (*d* = .35, favoring Whites). A recent report on the SRT detailed that an effect size in the .30s is not unusual for the measure and that an effect size of that magnitude is on par with both laboratory and applied research (Yusko et al., 2010).

Construct validity. The last couple of questions posed in the research were whether or not the neuropsychological and psychometric intelligence measures were related to each other and, if so, what is the magnitude of the relationships. The results indicated that all three intelligence tests were related to each other at a level of statistical significance ($p < .01$). Both psychometric intelligence measures were positively related to the neuropsychological intelligence battery at moderate levels ($r = .49$ to .51). The moderate correlations among the intelligence measures were beneficial in establishing construct validity for both psychometric and neuropsychological measures. The correlations were not so high (e.g., $r = .80$ and above), that they would indicate the tests measured the exact same aspects of the construct, but high enough to suggest that performance on the neuropsychological battery was related to the psychometric battery and the alternative psychometric assessment. That is, the neurobiological mechanisms needed to perform well on one measure are similar, or overlap, with the mechanisms needed to do well on another measure. In other words, they all asses the same construct, but in slightly different ways.

Prior to including neurological measures in with psychometric measures, psychometric assessments established construct validity by noting correlations with other psychometric measures (i.e., positive manifold). The Wonderlic, the Wechsler Adult Intelligence Scale (WAIS), or the Ravens Progressive Matrices can be considered gold standards of psychometric intelligence tests. New psychometric measures would establish construct validity with correlations to any one of the gold standard measures, but there was no link back to the brain. The current research, as well as another study that used psychometric assessments (Higgins et al., 2007), provide some of the evidence needed to bolster construct validity. There is empirical support that the neuropsychological intelligence battery is related to activation in the prefrontal cortex of the frontal lobes in the brain (Kane & Engle, 2002; Prabhakaran et al., 2000). The links between the neuropsychological battery and the other intelligence measures were the strongest in the current study (r = .49 to .51) and slightly weaker in Higgins and colleagues' study (i.e., correlations between the neuropsychological battery and the Wechsler Adult Intelligence Scale – Revised, a measure of psychometric intelligence, ranged from .20 to .45). Based on the results of the test of dependent correlations (i.e., hypothesis 10) there was no statistically significant difference between the neuropsychological battery’s relationship with the traditional psychometric battery and with the alternative psychometric assessment therefore the neuropsychological battery can be considered equally useful in establishing validity for both the psychometric battery and the alternative psychometric assessment.

Theoretical and Practical Implications

The current study is contributing to a small, but growing, body of literature that supports neuropsychological approach to intelligence as a predictor of performance. Compared to previous findings in which both neuropsychological and psychometric intelligence predicted academic performance, it was surprising that the neuropsychological intelligence battery was the only significant predictor of GPA. One reason why the current research may not have produced the expected predictive relationships is because of the sample drawn for the study. The Baruch College student body is comprised of non-traditional, often first-generation college students, and it is very racioethnically diverse. There is evidence that traditional psychometric assessments are culturally-bound and that these factors can impact performance (Fagan & Holland, 2007; Helms-Lorenz et al., 2003). The tasks involved in the neuropsychological battery are not culturally-bound, with the possible exception being the word fluency (i.e., word generation) task. The cultural aspects of the psychometric measures may have impacted performance to the degree that predictive relationships could not emerge within the data with this particular sample. One potential benefit of using a neuropsychological measure with a racioethnically diverse sample is that may be more predictive and culture-free than the traditional psychometric measures. The most culturally-bound assessment in the study was the traditional psychometric intelligence battery, and this was the only assessment that showed mean score differences between White/Caucasian American nationals and White/Caucasians born outside the United States. Clearly, culture impacted the current study and it should be considered as an influencing component in a model of intelligence.

The psychometric approach to intelligence has a long history as the dominant approach to intelligence research (Neisser et al., 1996). However, new information is emerging which necessitates a reevaluation of what should be included in a theoretical model of intelligence. As others have noted before (e.g., Deary, 2001), the definition and operationalization of intelligence needs to be expanded. These results demonstrate the value of incorporating neuropsychological aspects of intelligence within a new intelligence model. Neuropsychological intelligence has shown to predict performance but there is not enough data utilizing both approaches to intelligence to say definitively which approach is better. Perhaps it is not a question of which approach is better but rather multiple approaches are important. Regardless, additional research is needed that further explores the connections between neuroimaging and intelligence assessments, both psychometric and neuropsychological in nature.

Broad conclusions about the benefits of incorporating the neuropsychological approach into intelligence models cannot be made at this time. More data are needed to further understand the theoretical and practical outcomes. The Higgins study (Higgins et al., 2007) and the current study, both using neuropsychological and psychometric batteries, supported the finding that neuropsychological intelligence is a predictor of performance. However, the current study is the only research to investigate subgroup mean differences and it is unclear what the expected impact is on different racioethnic groups or what a “typical” effect size is for neuropsychological measures. Research has shown that the construct measured, the setting, and the sample impact the magnitude of mean score differences (Hough et al., 2001), therefore it is possible that the sample influenced the results of the current study. The current study should not be dismissed because of the unusual findings or lack of normalcy in some of the variables. These concerns are clearly outlined and results should be interpreted with caution. However, because certain relationships were not found in the current study that are supported in the literature (e.g., psychometric intelligence predicting performance, conscientiousness predicting performance), generalizing the results also should be done with caution.

Additionally more research is needed to see how the neuropsychological approach to intelligence relates to actual job performance (i.e., only supervisor ratings were used in prior research as a proxy for job performance) and job training performance. Information processing may be a stronger attribute of job and training performance compared to academic performance which may strengthen the argument for using neuropsychological measures. Regardless, there is more to learn about neuropsychological measures before taking them into an applied setting.

Industrial/organizational psychologists have the difficult task of determining the perfect balance between predictive validity of a tool and fair treatment of applicants (i.e., avoiding disparate outcomes for different gender or racioethnic groups). The neuropsychological battery has drastically reduced Black/White subgroup mean differences (d = .35) compared to traditional psychometric meta-analytic findings (d = 1.0; Hough et al., 2001). However, practically speaking, subgroup differences even of .30 may lead to adverse impact. Additionally, a major concern of the neuropsychological battery would be the disparate outcomes of males and females. On traditional tests of *g*, there are no reported mean differences for males and females, but using the neuropsychological battery may result in adverse impact on female applicants. While theoretical support for the neuropsychological approach to intelligence has been demonstrated, we now know that group differences do exist and could impact selection ratios, even if the differences are smaller compared to traditional psychometric intelligence measures.

Another practical concern of the neuropsychological battery is that it is very lengthy in its current form. The battery is typically a 90-minute administration but it can run longer. As the saying goes, “time is money” and human resources professionals value short and valid selection assessments. There is great promise in including measures of neuropsychological intelligence in predictive models or selection batteries, however improvements can and should be made to the battery before widespread implementation.

Limitations of the Current Study

There are a few limitations that should be taken into consideration when interpreting the results. One weakness of the current study is that there is only one criterion (i.e., self-report college GPA). The study design was limited to one criterion because of the time constraint of adding other criteria. Although, GPA is a very common criterion and used frequently in laboratory research therefore like comparisons with other research can be made. The fact that GPA was self-report was not perceived as a shortcoming because previous research has noted a high correlation ($r = .88$, $p < .001$) between transcript GPA and self-report GPA (Higgins et al., 2007) therefore it was selected as the criterion believing it to be an accurate proxy of academic performance. In retrospect, GPA may not have been the best criterion partially due to the high number of transfer students Baruch College receives every year. Transfer students’ previous schools may have graded very differently than Baruch College therefore it may not be an accurate reflection of a student’s achievement in comparison to students who have not transferred. It also may not be an accurate proxy of the student’s work if reported GPA was based on only one semester’s grades (in the case of participants in their first year of college coursework).

A suggestion for future research is to add other performance criteria. Having multiple criteria would be beneficial to further understand if various types of intelligence measures differentially predict performance outcomes. Some performance criteria considered for the present study were course grades (instead of cumulative GPA), change scores from midterm grades to final grades, to track GPA longitudinally, professor evaluations of the participant, evaluations from the participant's academic advisor, and supervisor performance ratings (if participant indicated he/she was currently working). Another performance criterion considered for the study was a computerized version of a cognitive task such as the Tower of Hanoi. The Tower of Hanoi task consists of three rods and four discs of ascending size conically stacked on the first rod. The objective is to move all the discs, one at a time, to different rods. The puzzle is solved correctly if all the discs end up on a different rod in the same (ascending) order as how the appeared when the task began. The Tower of Hanoi is regularly used in puzzle and adventure games. Ultimately, the ideas for more criteria had to be excluded from the study design due to time constraints or the feasibility of gaining access to accurate data (e.g., in the case of midterm and final grades).

A second possible limitation of the study may be test fatigue or lack of motivation. The neuropsychological battery lasts about 90 minutes and the entire study could take up to 3 hours to complete. The researcher acknowledges that if the study could have been shorter, then it would have been more pleasing to the participants, however the length of the measures did not appear to result in score differences. As previously noted, there were no score differences on any of the cognitive measures based on which length of measure occurred first (i.e., the 90-minute neuropsychological battery or the shorter 30-minute psychometric battery). Participants who were more than halfway through into the study scored no differently on the traditional psychometric intelligence battery compared to those who started the study with the psychometric battery. It is possible that fatigue or a drop in motivation may have set in on the last segment of the study. The alternative psychometric intelligence assessment which was taken last had a slightly lower average score (67%) compared to the traditional psychometric battery (69%) and had a larger standard deviation, relatively speaking. A second suggestion for future research was noted earlier in that it is important to reduce the length of the neuropsychological battery. The time of administration and the effect sizes may also decrease by removing some subdimensions. Obviously the validity of the new shorter battery will need to be tested.

A third possible limitation of the current study is that the study occurred in a laboratory setting. This type of setting lacks the realism of an applied setting. To stress the anonymity of the study, the participants read in the consent form and were also told by the research facilitator that performance on the study could not be matched backed to an individual. Participants were also told that their performance on the assessments had no bearing on their college course grade. After so much emphasis on the lack of accountability, one can question how motivated a study participant would be to perform and his or her best. Certainly, the level of motivation for the study participants was less than a hypothetical applicant pool taking the same assessments in hopes of gaining employment. However, a laboratory setting was appropriate for this research because no prior information on subgroup mean differences or adverse impact was available. Additionally, it would have been difficult or impossible to partner with an organization that was willing to use a 90-minute cognitive module in their selection system. Although, the study lacks external validity, the laboratory setting was a suitable starting point to help understand the implications of using a neuropsychological battery in a predictive model of performance. For future research, a shorter neuropsychological intelligence battery will make it easier to gain access to an organizational setting. It would be interesting to know what differences, if any, appear with employee and applicant samples.

Appendix A: Test Attitude Survey-Short

Please read each item carefully and use the response scale provided below.

1. Strongly Disagree
2. Disagree
3. Somewhat Disagree
4. Neutral
5. Somewhat Agree
6. Agree
7. Strongly Agree

Motivation

Doing well on this test (or these tests) is important to me.
I wanted to do well on this test or tests.
I tried my best on this test or tests.
I tried to do the very best I could to on this test or tests.
While taking this test or tests, I concentrated and tried to do well.
I want to be among the top scorers on this test (or these tests).
I pushed myself to work hard on this test or these tests.
I was extremely motivated to do well on this test or tests.
I just didn't care how I did on this test or tests. (R)
I didn't put much effort into this test or tests. (R)

Lack of Concentration

It was hard to keep my mind on this test or tests.
I found myself losing interest and not paying attention to the test or tests.
During the test session, I was bored.
I get distracted when taking tests of this type.

Belief in Tests

This test or tests was a good reflection of what a person could do in the job. (R)
Tests are a good way of selecting people into jobs. (R)
This kind of test or tests should be eliminated.
I don't believe that tests are valid.

Test Ease

This test was (or these tests were) too easy for me.
I found this test or tests too simple.
I found this test or tests interesting and challenging. (R)
I felt frustrated because many of the test questions were too difficult. (R)

External Attribution

I became fatigued and tired during the testing.
The questions on this test or tests were ambiguous and unclear.
I have not been feeling well lately and this affected my performance on the test or tests.
While taking the test or tests, I was preoccupied with how much time I had left.
I felt a lot of time pressure when taking this test or tests.

Appendix B: Demographic Questionnaire

1. Were you born in the U.S.?

    A. Yes
    B. No

2. Is English your first language?

    A. Yes
    B. No

3. Was English the primary language spoken in your household while you were growing up?

    A. Yes
    B. No

4. What is your current (college) GPA?

    A. Below 1.0
    B. 1.0 to 1.49
    C. 1.5 to 1.99
    D. 2.0 to 2.49
    E. 2.5 to 2.99
    F. 3.0 to 3.49
    G. 3.5 to 4.0

5. What is your SAT or ACT VERBAL score? If you took both tests only give an answer for one test.

    A. 200-300 SAT or 5-10 ACT
    B. 310-400 SAT or 11-16 ACT
    C. 410-500 SAT or 17-21 ACT
    D. 510-600 SAT or 22-26 ACT
    E. 610-700 SAT or 27-31 ACT
    F. 710-800 SAT or 32-36 ACT

Continue on to the next page ⇨

6. What is your SAT or ACT MATH score? If you took both tests only give an answer for one test.

    A. 200-300 SAT or 5-10 ACT
    B. 310-400 SAT or 11-16 ACT
    C. 410-500 SAT or 17-21 ACT
    D. 510-600 SAT or 22-26 ACT
    E. 610-700 SAT or 27-31 ACT
    F. 710-800 SAT or 32-36 ACT

7. What is your family's annual gross (pre-tax) income?

    A. $1,000 to $10,000
    B. $10,001 to $20,000
    C. $20,001 to $30,000
    D. $30,001 to $40,000
    E. $40,001 to $50,000
    F. $50,001 to $75,000
    G. $75,001 to $100,000
    H. $100,001 to $150,000
    I. More than $150,000

8. What was your high school GPA?

    A. Below 1.0
    B. 1.0 to 1.49
    C. 1.5 to 1.99
    D. 2.0 to 2.49
    E. 2.5 to 2.99
    F. 3.0 to 3.49
    G. 3.5 to 4.0

9. Before today, have you completed anything similar to the computerized cognitive assessment?

    A. Yes
    B. No

Table 1

Descriptive Statistics for Predictor and Criterion Variables

	N	•	*Min*	*Max*	*M*	*SD*
College Grade Point Average	292	-	1.00	7.00	5.84	1.04
Neuropsychological battery	247	-	-1.40	1.16	-0.34	0.44
Neuro battery first	161	-	-1.40	1.16	-0.33	0.44
Trad battery first	86	-	-1.31	0.96	-0.35	0.46
Trad psychometric battery	311	0.69	4.50	18.00	12.74	2.35
Neuro battery first	197	-	5.00	18.00	12.77	2.37
Trad battery first	113	-	4.50	17.50	12.70	2.34
Alt psychometric assessment	306	0.74	4.00	24.00	16.67	3.98
Neuro battery first	192	-	4.00	24.00	16.73	3.97
Trad battery first	113	-	6.00	24.00	16.62	3.99
Conscientiousness	247	-	-2.20	2.60	0.33	0.93

Note: Estimates of internal consistency were not available on neuropsychological battery and the conscientiousness personality factor because only dimension level scores were reported back from vendor. The report of “Neuro battery first” and “Trad battery first” refers to results from the tests of order effects.

Table 2

Intercorrelations of Predictor and Criterion Measures

	Correlations in Bold (*Sample sizes in Italics)*								
	1	2	3	4	5	6	7	8	9
1. College GPA	*292*	*233*	*292*	*287*	*233*	*233*	*233*	*233*	*233*
2. Neuro battery	**.14***	*247*	*247*	*242*	*247*	*247*	*247*	*247*	*247*
3. Trad psy battery	**.11**	**.49****	*311*	*306*	*247*	*247*	*247*	*247*	*247*
4. Alt psy assessment	**.09**	**.51****	**.48****	306	*242*	*242*	*242*	*242*	*242*
5. Conscientiousness	**.05**	**-.06**	**-.06**	**-.05**	*247*	*247*	*247*	*247*	*247*
6. Agreeableness	**.02**	**.09**	**.01**	**.08**	**.63****	*247*	*247*	*247*	*247*
7. Emotional Stability	**.04**	**.01**	**.02**	**-.01**	**.59****	**.60****	*247*	*247*	*247*
8. Extraversion	**.02**	**.06**	**.09**	**-.08**	**.34****	**.26****	**.46****	*247*	*247*
9. Openness/Intellect	**.01**	**.09**	**.11**	**.08**	**.49****	**.49****	**.50****	**.51****	*247*

* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table 3

Tests of Normality for Predictor Variables

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Neuro battery	.035	247	.200*	.035	247	.490
Trad psy battery	.078	311	.000	.078	311	.001
Alt psy assessment	.131	306	.000	.131	306	.000
Conscientiousness	.047	247	.200*	.047	247	.044

Notes. a. Lilliefors Significance Correction. * This is a lower bound of the true significance.

Table 4

Regression Models Predicting Academic Performance (GPA)

Hyp	Var	B	SE B	•	*t*	*p*
1a	Constant	5.95	0.085		69.94	0.00
	Neuro	0.33	0.153	0.14	2.18	0.03
1b	Constant	5.21	0.333		15.63	0.00
	Trad Psy	0.05	0.026	0.11	1.91	0.06
1c	Constant	5.49	0.264		20.65	0.00
	Alt Psy	0.02	0.015	0.09	1.52	0.13
2a	Constant	5.82	0.072		80.61	0.00
	Per-Cons	0.05	0.073	0.05	0.70	0.49
2b	Constant	5.94	0.088		67.65	0.00
	Neuro	0.34	0.153	0.15	2.22	0.03
	Per-Cons	0.06	0.073	0.05	0.83	0.41
2c	Constant	5.17	0.377		13.73	0.00
	Trad Psy	0.05	0.029	0.12	1.75	0.81
	Per-Cons	0.06	0.073	0.05	0.81	0.41

Table 5

Subgroup Mean Differences and Effect Sizes (d)

			Sex			Racioethnic				
		All	M	F	M/F *d*	B/AA	B/AA *d*	H	H *d*	W/C
GPA	*M*	5.84	5.95	5.76	-	5.46	-	5.67	-	6.06
	SD	1.04	0.92	1.12		1.11		1.09		0.93
	N	292	131	160		48		92		152
Neuro	*M*	-0.34	-0.28	-0.39	0.25	-0.45	0.33	-0.34	0.08	-0.30
	SD	0.44	0.45	0.43		0.46		0.47		0.42
	N	247	109	137		40		80		127
Trad Psy	*M*	12.74	13.18	12.39	0.34	12.31	0.21	12.81	0.01	12.83
	SD	2.35	2.20	2.42		2.19		2.05		2.56
	N	311	136	174		52		97		162
Alt Psy	*M*	16.67	16.87	16.50	0.09	15.43	0.35	17.03	-0.04	16.86
	SD	3.98	4.14	3.86		3.82		3.80		4.08
	N	306	134	171		51		96		159
CAL	*M*	-0.29	-0.17	-0.39	0.36	-0.44	0.31	-0.29	0.08	-0.24
	SD	0.63	0.61	0.64		0.75		0.63		0.59
WM	*M*	-0.43	-0.39	-0.45	0.12	-0.51	0.21	-0.45	0.10	-0.39
	SD	0.57	0.60	0.55		0.61		0.63		0.53
WFT	*M*	-0.58	-0.51	-0.64	0.15	-0.58	-0.01	-0.56	-0.03	-0.59
	SD	0.83	0.89	0.78		0.70		0.83		0.87

Table 5 *continued*

Subgroup Mean Differences and Effect Sizes (d)

Measure		All	Sex			Racioethnic				
			M	F	M/F *d*	B/AA	B/W *d*	H	H/W *d*	W/C
ANAT	*M*	-0.27	-0.18	-0.35	0.21	-0.45	0.33	-0.30	0.14	-0.20
	SD	0.78	0.78	0.78		0.85		0.82		0.73
ASAT	*M*	-0.26	-0.14	-0.35	0.24	-0.35	0.13	-0.24	0.01	-0.24
	SD	0.90	0.94	0.86		1.0		0.92		0.86
GNGT	*M*	-0.34	-0.17	-0.46	0.29	-0.51	0.21	-0.32	0.03	-0.29
	SD	1.03	0.95	1.07		1.09		0.99		1.03
ROST	*M*	-0.72	-0.58	-0.82	0.25	-0.84	0.23	-0.79	0.17	-0.63
	SD	0.94	0.95	0.93		0.97		1.0		0.90
RLST	*M*	-0.37	-0.38	-0.36	-0.02	-.37	-0.05	-0.31	-0.13	-0.42
	SD	0.83	0.89	0.77		0.78		0.84		0.84
RJT	*M*	-0.19	-0.20	-0.18	-0.02	-0.31	0.21	-0.24	0.12	-0.12
	SD	0.94	0.99	0.90		0.94		1.0		0.90

Notes: Selected subdimensions of the neuropsychological battery are conditional associative learning (CAL), working memory (WM), word fluency task (WFT), conditional non-spatial association test (ANAT), conditional spatial association test (ASAT), go no-go task (GNGT), random object span test (ROST), and recency judgment task (RJT). Negative *d* coefficients indicate the test favors the minority group (i.e., females, Black/African Americans, Hispanics).

Table 6

Descriptive Statistics for Predictor Variables of White/Caucasian Participants

		N	*Min*	*Max*	*M*	*SD*	*d*	*t*	*p*
1. Were you born in the U.S.?									
Neuro	Yes	60	-1.24	0.63	-0.24	0.44	0.31	1.73	0.086
	No	66	-1.40	0.96	-0.37	0.39			
Trad Psy	Yes	72	8.50	18.00	13.74	2.21	0.68	4.29	0.000
	No	88	4.50	17.50	12.07	2.62			
Trad Psy	Yes	71	4.00	24.00	17.14	4.43	0.13	80	0.426
	No	86	5.00	23.00	16.62	3.81			
2. Is English your first language?									
Neuro	Yes	62	-1.24	0.63	-0.27	0.43	0.17	0.97	0.334
	No	64	-1.40	0.96	-0.34	0.41			
Trad Psy	Yes	76	8.50	18.00	13.76	2.17	0.73	4.62	0.000
	No	84	4.50	17.50	11.98	2.64			
Trad Psy	Yes	75	4.00	24.00	16.96	4.52	0.05	0.31	0.756
	No	82	7.00	23.00	16.76	3.69			

Table 6 *continued*

Descriptive Statistics for Predictor Variables of White/Caucasian Participants

		N	*Min*	*Max*	*M*	*SD*	*d*	*t*	*p*
3. Was English the primary language spoken in your household while you were growing up?									
Neuro	Yes	62	-1.24	0.63	-0.27	0.43	-0.03	-0.14	0.889
	No	64	-1.40	0.96	-0.34	0.41			
Trad Psy	Yes	76	8.50	18.00	13.76	2.17	0.51	3.07	0.003
	No	84	4.50	17.50	11.98	2.64			
Trad Psy	Yes	75	4.00	24.00	16.96	4.52	-0.15	-0.92	0.402
	No	82	7.00	23.00	16.76	3.69			

Table 7

Percentage of Variance Explained in the Academic Performance Measures by the Cognitive Ability Measures

	Cognitive ability measures		
Performance data	Neuro	Trad Psy	Alt Psy
College GPA	95%	12%	9%
Log Transform College GPA	98%	10%	5%

References

Arvey, R. D., Strickland, W., Drauden, G., & Martin. C. (1990). Motivational components of test-taking. *Personnel Psychology,* 43, 695-716.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114,* 542-551.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Thousand Oaks: Sage.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687-732). Palo Alto, CA: Consulting Psychologists Press, Inc.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

Cattell, R. B. (1973). *Measuring intelligence with the culture fair tests.* The Institute for Personality and Ability Testing, Champaign, IL.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. New York: Elsevier Science Publishing Co.

Cianciolo, A. T., & Sternberg, R. J. (2004). *Intelligence: A brief history.* Malden, MA: Blackwell Publishing.

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

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7,* 547-552.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

Cross, B., & Travaglione, A. (2003). The untold story: Is the entrepreneur of the 21st century defined by emotional intelligence. *The International Journal of Organizational Analysis, 11*, 221-228.

Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental-differential approach. *Trends in Cognitive Sciences, 5,* 164-170.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.) *The cognitive neurosciences* (pp. 721-733). Cambridge, MA: MIT Press.

Duncan, J. (2005). Frontal lobe function and general intelligence: Why it matters. *Cortex, 41*, 215-217.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33,* 261-268.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179-197.

Fagan, J. F. (2000). A theory of intelligence as processing: Implications for Society.

Psychology, Public Policy, and Law, 6, 168-179.

Fagan, J. F., & Holland, C. R. (2002). Equal opportunity and racial differences in IQ. *Intelligence, 30,* 361-387.

Fagan, J. F., & Holland, C. R. (2007). Racial equality in intelligence: Predictions from a theory of intelligence as processing. *Intelligence, 35,* 319-334.

Ferreter, J. M., Goldstein, H. W., Scherbaum, C. A., Yusko, K. P., & Jun, H. (2008, April). *Reducing adverse impact using a nontraditional cognitive ability assessment*. Poster presented at the 23rd annual conference of the Society for Industrial and Organizational Psychology, San Francisco.

Frith, C. D., Friston, K. J., Liddle, P. F., & Frackowiak, R. S. (1991). A PET study of word finding. *Neuropsychologia, 29,* 1137-1148.

Freedle, R. (1986). *Second interim report: Report on results of a semantic scoring of four analogy subtest including preliminary findings of two projects dealing with application of a statistic for comparing Black and White examinees on SAT analogies and GRE analogies.* Unpublished report, dated May 16, 1986, Educational Testing Service, Princeton, N. J.

Freedle, R., & Kostin, I. (1988). *Relationship between item characteristics and an index of differential item functioning (DIF) for the four GRE verbal item types.* Research Report RR-88-29, Educational Testing Service, Princeton, N. J.

Freedle, R., & Kostin, I. (1990). Item difficulty of four verbal item types and an index of differential item functioning for Black and White examinees. *Journal of Educational Measurement, 27,* 329-343.

Freedle, R., & Kostin, I. (1997). Predicting Black and White differential item functioning in verbal analogy performance. *Intelligence, 24,* 417-444.

Gabel, R. S., Dolan S. L., & Cerdin J. L. (2005). Emotional intelligence as predictor of cultural adjustment for success in global assignments. *Career Development International, 10,* 375-379.

Galton, F. (1883). *Inquiry into Human Faculty and its Development*. London: Macmillan.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in the neural plasticity as an explanatory mechanism. *Psychological Review, 109,* 116-136.

Gibson,W. A. (1962). Orthogonal predictors: A possible resolution of the Hoffman-Ward controversy. *Psychological Reports, 11,* 32-34.

Grigorenko, E. L. (2000). Heritability and intelligence. In R. J. Sternberg, (Ed.), *Handbook of intelligence* (pp. 53-91). New York: Cambridge University Press.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4,* 26-42.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84-96.

Goldstein, H. W., Scherbaum, C. A., & Yusko, K. P. (2009). Revisiting *g*: Intelligence, adverse impact, and personnel selection. In J. Outtz, (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 95-134). Mahwah, NJ: Lawrence Erlbaum.

Goldstein, H. W., Zedeck, S., & Goldstein, I. L. (2002). *g*: Is this your final answer?

Human Performance, 15, 123-142.

Goleman, D. (1995). Emotional intelligence. New York: Bantam Books.

Gould, S. J. (1995). Curveball. In S. Fraser (Ed.), *The bell curve wars: Race, intelligence, and the future of America* (pp.11-22). New York: Basic Books.

Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: Science and ethics. *Nature, 5,* 471-482.

Haier, R. J. (2003). Brain imaging studies of intelligence: Individual differences and neurobiology. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of intelligence: International perspectives.* Washington, DC: American Psychological Association.

Haier, R. J., Siegel Jr., B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: A positron emission tomographic study. *Brain Research, 570,* 134-143.

Helms-Lorenz, M., Van de Vijver, F. J. R., & Poortinga, Y. H. (2003). Cross-cultural differences in cognitive performance and Spearman’s hypothesis: *g* or *c*? *Intelligence, 31*, 9-29.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Higgins, D. M., Peterson, J. B., Pihl, R. O., & Lee, A. G. M. (2007). Prefrontal cognitive ability, intelligence, big five personality, and the prediction of advanced academic and workplace performance. *Journal of Personality and Social Psychology, 93,* 298-319.

Hoffman, P. J. (1962). Assessment of the independent contributions of predictors. *Psychological Bulletin*, *59,* 77-80.

Horton, A. M., Jr., & Wedding, D. (1984). *Clinical and Behavioral Neuropsychology*. New York: Praeger.

Hough, L. M., Oswald, F. L., Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96,* 72-98.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85,* 869-879.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R. (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence, 17,* 44-77.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35,* 1-19.

Johnson, J. W. (2001). Determining the relative importance of predictors in multiple regression: Practical applications of relative weights. *Advances in Psychological Research, 5,* 231-251.

Johnson, J. W. (2004). Factors affecting relative weights: The influence of sampling and measurement error. *Organizational Research Methods, 7,* 283-299.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9,* 637-671.

Lezak, M. D. (1983). *Neuropsychological assessment.* New York: Oxford University Press.

Linderman, R. H., Merenda, P. F., & Gold, R. A. (1980). *Introduction to bivariate and multivariate analysis.* Glenview, IL: Scott, Foresman and Company.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.) *Handbook of intelligence* (pp. 176-193). New York: Cambridge University Press.

Lukin, M. E., Dowd, T., Plake, B. S., & Kraft, R. G. (1985). Comparing computerized versus traditional psychological assessment. *Computers in Human behavior, 1,* 49-58.

McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence, 33,* 337-346.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114,* 449-458.

Meier, M. J. (1974). Some challenges for clinical neuropsychology. In R. M. Reitan & L. A. Davison (Eds.), *Clinical neuropsychology: Current status and applications* (pp. 289-323). Washington, DC: V. H. Winston and Sons.

Milner, B., & Petrides, M. (1984). Behavioural effects of frontal-lobe lesions in man. *Trends in Neurosciences, 7,* 403-407.

Milner, B., Petrides, M., & Smith, M. L. (1985). Frontal lobes and the temporal organization of memory. *Human Neurobiology, 4,* 137-142.

Moore, E. G. J. (1986). Family socialization and the IQ test performance of traditionally and transracially adopted black children. *Developmental Psychology, 22,* 317-326.

Mount, M. K., Oh, I., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology, 61*, 113-139.

Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than *g*. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3-30). San Francisco: Jossey-Bass Publishers.

Murray, C., & Herrnstein, R. J. (1994). Race, genes and I.Q. – an apologia. *The New Republic*. October 31.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.

Nisbett, R. E. (2005). Heredity, environment, and race differences in IQ: A commentary on Ruston and Jensen (2005). *Psychology, Public Policy, and Law, 11*, 302-310.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics, 51*, 1352-1375.

Oswald, F. L., Johnson, J. W., & Oliver, D. H. (2000, April). The importance of relative importance weights: Statistical and rational considerations. In J. W. Johnson (Chair), *Practical applications of relative importance methodology in I/O psychology*. Symposium conducted at the 15th Annual Conference of the Society for Industrial and Organizational Psychology. New Orleans, LA.

Petrides, M. (1985). Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia, 23,* 601-614.

Petrides, M. (1990). Nonspatial conditional learning impaired in patients with unilateral frontal but not unilateral temporal lobe excisions. *Neuropsychologia, 28,* 137-149.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia, 20,* 249-262.

Petrides, M., Alivisatos, B., Evans, A. C., & Meyer, E. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academy of Sciences, USA, 90,* 878-882.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61,* 153-172.

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience, 3,* 85-90.

Pyburn, K., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61,* 143-151

Raju, N. S., Drasgow, F., & Slinde, J. A. (1993). An empirical comparison of the area methods, Lord's chi-square test and the Mantel-Haenszel technique for assessing differential item functioning. *Educational and Psychological Measurement, 53,* 301-314.

Richardson, K. (2005). *g* is about variance not contents. *Cortex, 41,* 241-242.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297-330.

Rushton, J. P. (1994). Sex and race differences in cranial capacity from international labour office data. *Intelligence, 19*, 281-294.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11*, 235-294.

Sackett, P. R., and Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*, 929-954.

Scarr, S., & Weinberg, R. (1976). IQ test performance of black children adopted by white families. *American Psychologist, 31,* 726-739.

Scherbaum, C. S., & Goldstein, H. W. (2008). Examining the relationship between differential item functioning and item difficulty. *Educational and Psychological Measurement, 68*, 537-553 .

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Shallice, T., & Burgess, P. W. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds), *Frontal lobe function and dysfunction* (pp. 125-138). Oxford: Oxford University Press.

Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Spearman, C. E. (1927). *The abilities of man, their nature and measurement.* New York: Macmillan.

Sternberg, R. J. (1981). Intelligence and non-entrenchment. *Journal of Educational Psychology, 73*, 1-16.

Sternberg, R. J. (1984). A contextualist view of the nature of intelligence. In P.S. Fry (Ed.), *Changing conceptions of intelligence and intellectual functioning: Current theory and research.* New York: North-Holland.

Sternberg, R. J. (2005). The importance of converging operations in the study of human intelligence. *Cortex, 41,* 243-244.

Sternberg, R. J. (2006). The rainbow project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence, 34,* 321-350.

Sternberg, R. J. & Detterman, D. K. (Eds.) (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 155-162). Norword, NJ: Ablex.

Teasdale, T. W., & Pakkenberg, B. (1988). The association between intelligence level and brain volume measures: A negative finding. *Scandinavian Journal of Psychology, 29*, 123-125.

Van der Maas, H., Dolan, C., Grasman, R., Wicherts, J., Huizenga, H., & Raijmakers, M. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113,* 842-861.

Wiesen, J. P. (2007). Possible reasons for the black-white mean score differences seen with many cognitive ability tests: Informal notes to file. Retrieved November 18, 2008 from http://appliedpersonnelresearch.com/papers/adimpact.pdf

Williams, W. M. (1996). Consequences of how we define intelligence. *Psychology, Public Policy, and Law, 2*, 506-535.

Williams, W. M., & Ceci, S. J. (1997). Are Americans becoming more or less alike? Trends in race, class, and ability differences in intelligence. *American Psychologist, 52*, 1226-1235.

Yusko, K., Goldstein, H. G., Oliver, L. O., Hanges, P. J. (2010, April). Cognitive Ability Testing With Reduced Adverse Impact: Controlling for Knowledge. In C. Paullin (Chair), *Cognitive ability testing: Exploring new models, methods, and statistical techniques*, symposium conducted at the 25th Annual Conference of the Society for Industrial and Organizational Psychology. Atlanta, GA.