

STEREOTYPE THREAT THEORY AS AN  
EXPLANATION FOR THE DEPRESSED  
PERFORMANCE ON COGNITIVE ABILITY  
MEASURES BY AFRICAN AMERICANS

By

JOHN MICHAEL NOMURA

Bachelor of Arts  
University of Oklahoma  
Norman, Oklahoma  
2000

Master of Science  
Oklahoma State University  
Stillwater, Oklahoma  
2004

Submitted to the Faculty of the  
Graduate College of the  
Oklahoma State University  
in partial fulfillment of  
the requirements for  
the Degree of  
DOCTOR OF PHILOSOPHY  
July, 2007

STEREOTYPE THREAT THEORY AS AN  
EXPLANATION FOR THE DEPRESSED  
PERFORMANCE ON COGNITIVE ABILITY  
MEASURES BY AFRICAN AMERICANS

Report Approved:

---

Terry Stinnett, Ph.D.  
Dissertation Adviser

---

Dale Fuqua, Ph.D.

---

Gary Duhon, Ph.D.

---

John Chaney, Ph.D.

---

A. Gordon Emslie  
Dean of the Graduate College

COPYRIGHT

By

John Michael Nomura

July 2007

## ACKNOWLEDGEMENTS

I would like to thank my committee members for their generous help with this project. I am grateful for the academic freedom afforded to the students in our program. The many, disparate projects that result from such flexibility must be burdensome for the faculty. So, I am doubly grateful for their constant availability and encouragement. I believe that my trainers have endeavored to foster skepticism, critical thinking, and an appreciation for sound research design in their students. I hope that such virtues are manifest in this report. I would like to thank my chair, Terry Stinnett, Ph.D., for his assistance with the development of my ideas and his substantial help with the development of this study. I am also grateful to Judy Oehler-Stinnett, Ph.D., for the hours she spent listening to my ideas, providing alternative perspectives, and offering valuable constructive criticisms. I am indebted to Dale Fuqua, Ph.D. for providing additional perspective to my statistical interpretation of ST theory, and for examining my analyses. I would like to thank Gary Duhon, Ph.D. for encouraging me to follow my interests. I would like to thank my family for their support and encouragement. I would especially like to thank my wife who patiently tolerated scores of books, articles, and protocols cluttering our home for years. I would also like to thank her for her intellectual encouragement of my work. Without funding, a study of this scope would not have been possible without talented volunteers. I would like to thank the following research assistants: Michelle

Atkins, Shannon Beason, Felicia Castro, Chawndolyn Gore, Karen Hogan, Quinita Johnson, Sally Linden, Brandi Newry, Erika Olinger, and Kim Wiechmann. Like all my work, this is for my baby girl, Maisie.

Lastly, I would like to dedicate this to all my friends over the years who were proud to call themselves Black; especially Walter Morgan, Charles Rayford, and Terry Williams. It saddens me that they must live in a culture where prejudiced individuals incessantly underestimate their remarkable talents. I would also like to express my gratitude to Langston University, a traditionally Black institution that allowed me to recruit participants from among the students. The LU IRB was keenly aware that this was a critical review, yet they were confident enough to know that the truth about racial differences is on their side, and exposure to examination could only benefit them.

## PREFACE

I believe that my decision to examine the sensitive issue of racial differences in cognitive performance carries with it the obligation to disclose my motives. Because I have chosen to present a balanced review of current literature, some of my content may offend readers who subscribe to a particular perspective on this contentious topic. To be clear, my critique of stereotype threat theory is not an attempt to align myself with hereditarians. I would like to emphatically state that I harbor no prejudice toward any ethnic group. I believe that all students are equal and are capable of learning. I refuse to recognize “race” as a function for any physical, emotional, behavioral, or intellectual deficit. Nevertheless, Black students whom I serve disproportionately experience negative educational outcomes, and I feel that those of us who are in a position to improve this situation should do so. I hope that readers of this report will be empowered to more readily identify bias and inaccuracy in research on racial inequities, and consequently will be more effective in improving the lives of Black Americans.

Although I am deeply concerned about the continued underperformance of Black students, the impetus behind this project was primarily my fascination with bias in psychology research. I have always been a skeptic, and when I was first exposed to stereotype threat theory as a young graduate student, I quickly identified the erroneous statistical interpretation. I naively believed that I had

stumbled upon a remarkable phenomenon: a widespread misunderstanding of research design. Only in the intervening years have I realized that such fallacy is not all that exceptional in this field. From the homogenous theoretical orientations of journal editorial boards to the political agendas of grant committees, psychology theory in America is, like all sciences, predicated on the values of the theorists. I hope that this report exposes how personal bias can dominate objectivity to the detriment of reason and progress. My hope is that anyone who chooses to read this report will be more skeptical in the future- not of their colleagues, but of themselves. Every unchecked personal bias is an impediment to progress, and a thousand unchecked biases can move us wholly in the wrong direction.

## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION .....	1
Race .....	4
Cognitive Ability .....	5
Race/Cognitive Ability Relationship .....	6
Stereotype Threat .....	7
Research Question .....	8
II. LITERATURE REVIEW .....	9
Race .....	9
Definitions of Race .....	10
Racial Differences .....	12
Bias .....	15
Personal Bias .....	16
Bias and <i>The Bell Curve</i> .....	16
Bias and Afrocentrism .....	19
Bias and the Pioneer Fund .....	22
Bias Summary .....	24
Cognitive Ability .....	25
IQ .....	25
G factor .....	28
Black/White IQ Gap .....	30
Introduction to contemporary explanations .....	35
Hereditarian theories .....	38
Environmental deficit theories .....	41
Cultural relativism theories .....	45
Stereotype Threat .....	49
Criticisms of Stereotype Threat .....	57
Subsequent Research on Stereotype Threat .....	67
III METHODS .....	73
Participants .....	74
Procedure .....	75
Overview .....	75
Detailed Procedure .....	75



Task Description Manipulations .....	79
Instrument .....	83
Planned Analyses .....	86
Research Questions and Hypotheses .....	86
Research Question 1 .....	87
Research Question 2 .....	87
Research Question 3 .....	88
Additional Hypotheses .....	88
 IV RESULTS.....	 90
Sample .....	91
Assumptions.....	91
Analyses on Nonverbal Performance.....	92
Analyses on Verbal Performance.....	98
Analysis of Deception Check.....	103
Research Questions.....	104
Research Question 1 .....	104
Research Question 2 .....	105
Research Question 3 .....	105
Additional Hypotheses .....	107
Limitations and Future Research .....	107
 V DISCUSSION.....	 112
Mischaracterization .....	113
ST Predictions.....	116
The G Factor.....	117
ST Threatened Populations .....	118
Generalizability .....	120
New ST Effect.....	123
Uncompromised Psychometrics.....	123
Conclusion .....	124
New ST Phenomenon.....	123
Argumentum Ad Hominem.....	126
Reasonable Suspicion .....	128
Kant Quote.....	130
 REFERENCES .....	 131
 APPENDICES .....	 141
Appendix A: Cognitive Task Cover for the Control Condition .....	141
Appendix B: Cognitive Task Cover for the Stereotype Threat Condition.....	142

Appendix C: Cognitive Task Cover for the No Stereotype Threat Condition.....	142
Appendix D: OSU IRB approval letter.....	143
Appendix E: Langston University IRB approval letter .....	144
Appendix F: OSU University Registrar Approval Letter .....	145
Appendix G: Langston University Registrar Approval Letter.....	146
Appendix H: Response Sheet Front for the Control Condition .....	147
Appendix I: Response Sheet Back for the Control Condition .....	147
Appendix J: Response Sheet for the Stereotype Threat Condition .....	148
Appendix K: Response Sheet for the No Stereotype Threat Condition ....	149

## LIST OF TABLES

Table	Page
1. Raw Score to IQ Score Conversion for Brown and Day 2006 Article .....	71
2. Conformity of Data to Statistical Assumptions by Analysis .....	92
3. Descriptive Statistics for ANOVA on Nonverbal Performance .....	95
4. Analysis of Variance for Nonverbal Performance .....	95
5. Post-Hoc Analyses on Nonverbal Performance Data .....	96
6. Descriptive Statistics for ANCOVA on Nonverbal Performance .....	96
7. Analysis of Covariance for Nonverbal Performance .....	96
8. Descriptive Statistics for Analyses of Verbal Performance .....	101
9. Analyses of Verbal Performance .....	101
10. Percentage of Participants' Responses by Experimental Condition .....	104
11. Difference Between Means of Obtained Samples and National Scores by Study .....	122

## LIST OF FIGURES

Figure	Page
1. Research Hypotheses .....	89
2. Mean Performance by Racial Groups on Nonverbal Cognitive Task .....	97
3. Adjusted Mean Performance by Racial Groups on Nonverbal Cognitive Task .....	97
4. Mean Performance by Racial Groups on Verbal Cognitive Task .....	102
5. Adjusted Mean Performance by Racial Groups on Verbal Cognitive Task .....	102

## NOMENCLATURE

ANCOVA	analysis of covariance
ANOVA	analysis of variance
APM	Raven's Advanced Progressive Matrices
Cx	Control Condition
DC	diagnostic condition
Df	degrees of freedom
DV	dependent variable
g	general factor of intelligence
GLM	general linear model
GRE	graduate record examination
IQ	intelligence quotient (mean=100, SD=15)
IRB	Institutional Review Board
ISOT	Integrated Scholastic-Occupational Test
IV	independent variable
LU	Langston University
MMR	mild mental retardation
NDC	nondiagnostic condition
NDCC	nondiagnostic-challenge condition

NSTC	no stereotype threat condition
OSU	Oklahoma State University
p	probability, usually of error (p=.05 means 5% chance of error)
r	correlation coefficient
SAT	Originally, this stood for scholastic aptitude test. Currently, the letters have no official meaning.
SD	standard deviation
SES	socioeconomic status
SPSS	Statistical Package for the Social Sciences (software)
ST	stereotype threat
STC	stereotype threat condition
TD	Task Description
$\eta_p^2$	Partial Eta Squared

CHAPTER I  
INTRODUCTION

Stereotype threat theory, as it is known today, was first proposed by Claude Steele and Joshua Aronson in their 1995 collaborative article, *Stereotype threat and intellectual test performance of African Americans*. Through a series of five experiments, Steele and Aronson presented the stereotype threat phenomenon as a factor that they believed may explain a significant amount of the discrepancy in cognitive performance between Blacks and Whites in the US, especially performance on standardized tests. Since their seminal publication, scores of articles and several books have addressed the theory. Whereas Steele and Aronson's 1995 article investigated the depressed performance of Blacks, they and others have since researched the applicability of the phenomenon to other demographic groups, including women, Latinos, and Whites. Some studies have attempted to investigate mediating factors and underlying mechanisms. Also, some authors, including Aronson, have published literature that purports to instruct educators and interested persons in how the stereotype threat theory may be used to intervene with the depressed performance of affected demographic groups.

Remediating racial differences in cognitive performance and academic achievement is germane to the work of all school psychologists, especially if such differences are artifacts of testing procedures or other sources of bias. First, school psychologists are experts in cognitive assessment and academic achievement. They have a responsibility to contribute to current literature in that capacity. Secondly, many school psychologists are accountable for effecting positive outcomes for students of all ethnic backgrounds. Ensuring accurate assessments of cognitive ability and developing an accurate understanding of causal factors in academic achievement is paramount to that task.

Recently, there has been a growing sentiment within school psychology to move away from IQ testing and diagnostic labeling to a more applied, pragmatic approach (Kush et al., 2001). Proponents of this movement are interested in client-specific variables and may have less interest in stereotype threat research that examines group performance differences on IQ and aptitude tests. However, understanding nomothetic trends and underlying mechanisms in performance on such measures may facilitate greater efficacy in idiographic applications and client interventions.

Furthermore, performance on IQ and aptitude tests is directly relevant to school psychologists' work. It is currently the most reliable predictor of academic achievement (Braden, 1995; Sattler, 2001), is correlated with a variety of advantageous life outcomes (Herrnstein & Murray, 1994; Oakland, 1995), and is important to educational and social policy (Braden, 1995). Also, performance on IQ tests is routinely used by school psychologists to place students in special

education programs. Disadvantaged children, many of whom are placed in special education programs, command a majority of resources spent in education. For every \$56 spent on educational programs that benefit all children, \$922 are spent on disadvantaged youth and only \$1 is spent on gifted children (Braden, 1995; Herrnstein & Murray, 1994).

For school psychologists, one of the most salient issues in racial differences studies concerns the over-representation of minorities in special education. Blacks are about 2.5 times more likely to be identified as mildly mentally retarded (MMR) and 1.5 times more likely to be identified as seriously emotionally disturbed (SED) than Whites. Lower SES of the identified individuals and higher affluence of the school district are known to be positively correlated with increasing Black over-representation in MMR (Oswald, Coutinho, Best, & Singh, 1999). Whites are 3.2 times more likely to be assigned to a gifted class than Blacks. Conversely, Asian Americans are typically under-represented in special education disability categories and over-represented in classes for the gifted and talented (Artiles & Trent, 1994). The consistency with which Blacks exhibit a lower mean on cognitive ability tests (IQ tests) is critical to the problem of African-American over-representation in special education and inseparably related to the problematic racial IQ gap. Understanding factors such as the stereotype threat phenomenon, which may contribute to depressed academic performance of Blacks, is imminently important to society and must be addressed by school psychologists.



The purpose of this study is to investigate the effect of stereotype threat elicitors on the cognitive performance of African-Americans and whether the stereotype threat effect can account for the discrepancy in performance between African-American students and Caucasian-American students. Specifically, this study addresses the following question: Do stereotype threat elicitors differentially affect the performance of Blacks and Whites on standardized cognitive tests such that removal of stereotype threat elicitors will eliminate the discrepancy in cognitive performance between those groups? In addressing this question, several constructs (or variables) must first be discussed. These variables are (1) race, (2) cognitive ability, (3) the relationship between race and cognitive ability, and (4) stereotype threat. The remainder of this introduction discusses these variables via condensed excerpts from Chapter II. Interested readers are encouraged to skip to Chapter II to avoid reading redundant material.

## Race

Because classification is a result of subjective experience, it should come as no surprise that there is significant inconsistency in the manner by which races are categorized. Race classification has been based on language (Molnar, 1983), geography, physical characteristics, and even religion (Frisby, 1993a). One traditionally employed operational definition divides individuals into three distinct groups: Negroid (Blacks), Mongoloid (Asians), and Caucasoid (Whites) (Levin, 1997). Jensen (1998) reviewed a principal component analysis of 42 populations and suggested that two components, one related to geographic

migration distance and another related to climate, account for 43% of the genetic variance between racial groups. Other authors have argued that race is simply a social construct and any racial classification system has no true validity (Sternberg, 2005). According to Herrnstein and Murray (1994), most studies in racial differences are based on how subjects choose to classify themselves. Thus defined, racial groups have been shown to differentially vary across various physiological, psychological, and social constructs (see Herrnstein & Murray, 1994; Levin, 1997; Overfield, 1995; Molner, 1983; Weiss & Mann, 1981).

### Cognitive Ability

Cognitive ability, or intelligence, is measured by IQ tests. Such tests are not direct measures of a person's intelligence, but are estimates of an individual's general intellectual functioning. IQ scores are derived by comparing the individual's performance to the performance of people of the same age on various cognitive tasks. Historically, IQs are reported as standard scores that have a mean of 100 and a standard deviation of 15-16. They are the best predictors of academic achievement (Jensen, 1998; Oakland, 1995; Sattler, 2001) and are correlated with a variety of advantageous life outcomes (Herrnstein & Murray, 1994; Oakland, 1995). Factor analyses of IQ tests reveal a general factor of intelligence (the g factor) that accounts for variance across all types of IQ tests and subtests. There are several current theories that describe additional cognitive abilities that are more specific than the g-factor (Jensen, 1998). These factors include Fluid and Crystallized Intelligence. IQ is positively

correlated with most other tests of cognitive ability, including standardized tests of achievement and aptitude (Herrnstein & Murray, 1994). Stereotype threat research has been applied to tasks related to Fluid Intelligence, or Gf (McKay, 1999) and Crystallized Intelligence, or Gc (Steele & Aronson, 1995).

### Race/Cognitive Ability Relationship

It is widely believed that, regardless of the cause, there is approximately one standard deviation difference in mean performance between Blacks and Whites on IQ tests and other measures of cognitive ability (Frisby, 1995; Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). More specifically, Blacks exhibit a mean IQ of 85.

A preponderance of evidence shows that the depression of Blacks' IQ scores is due to factors other than systematic bias in IQ tests. The evidence includes the following trends: The IQ gap is exacerbated by culturally neutral and g-loaded items and tests; IQ tests do not under-predict and may over-predict Black performance on several criteria (academic achievement, college success, etc.); Blacks perform more poorly than Whites on test items that were chosen to maximize Blacks' performance; there is excellent factorial similarity between Black and White standardization samples; the discrepancy exists across many cultures/countries; test characteristics (other than g loadings) do not predict the discrepancy (see Herrnstein & Murray, 1994; Jensen, 1998; Kush et al., 2001; Levin, 1997; Noble, 2003). This research does not rule out environmental factors accounting for the IQ gap, but it suggests that the gap is not due to measurement

bias. Stereotype threat researchers propose that the gap is due to measurement bias.

### Stereotype Threat

The original authors offer the following definition of stereotype threat:

When a negative stereotype about a group that one is part of becomes relevant, usually as an interpretation of one's behavior or an experience one is having, stereotype threat is the resulting sense that one can then be judged or treated in terms of the stereotype or that one might do something that would inadvertently confirm it. (Steele, Spencer, & Aronson, 2002)

Generally, it is theorized that when an African American takes an IQ test, he/she may become cognizant of negative stereotypes about Black intellectual inferiority. Research has shown that these stereotypes remain widely known (for example, see Devine & Elliot, 1995). Stereotype threat infers that the subject, now aware of the stereotype and its relevance to a task which he/she is about to perform, becomes concerned about confirming the stereotype. It is suggested that this fear may be related to the possibility of confirming the stereotype to oneself or permitting others to confirm it through one's actions, thus facilitating the stereotype's perpetuation. This confirmation fear allegedly depresses the subject's performance.

Researchers use experimental manipulations to elicit stereotype threat. There are two types of elicitors commonly used. One elicitor is the description of the test as diagnostic of a person's ability. This is referred to as the Diagnosticity manipulation. The second is called the Race Priming manipulation. This elicitor

involves asking the subject for a self-report of race prior to a cognitive task (see Steele & Aronson, 1995). Each of these elicitors has been shown to significantly depress the performance of African Americans (Steele & Aronson, 1995; Claytie Davis III, 2000, Joseph Brown, 2001). Because stereotype threat researchers statistically remove variance in performance related to individual and group differences in cognitive ability prior to analysis of the stereotype threat effect, it remains unclear whether stereotype threat accounts for any portion of the racial performance gap (Sackett, Hardison, & Cullen 2004).

#### Research Question

The primary question to be answered by this study is whether stereotype threat elicitors differentially affect the performance of Blacks and Whites on standardized cognitive tests such that absence of stereotype threat elicitors will decrease the discrepancy in performance between those groups.

## CHAPTER II

### LITERATURE REVIEW

The purpose of this literature review is to provide a critical and contextual review of stereotype threat theory. Research will be discussed concerning all the major factors involved in stereotype threat theory. These factors include race, cognitive ability (IQ), the relationship between race and cognitive ability, and stereotype threat. Additionally, this chapter will draw on a diversity of sources to present stereotype threat theory in its academic and political context. The controversial nature of the subject of racial differences necessitates such a contextual understanding. Thus, other factors discussed will include sources of bias in racial research and contemporary explanations for racial differences in cognitive performance.

#### Race

In reviewing literature concerning racial differences, a number of issues must first be addressed. These issues include approaches to racial classification, validity of racial categories, and biases that pervade literature concerning racial differences. Political, social, cultural, historical, and legal factors interact to often polarize the issue of racial differences. The arguments

concerning the contentious subject of racial differences that are forwarded by authors may sometimes be influenced by the authors' motives and biases as much as the empirical evidence. This trend is clearly exhibited in stereotype threat research; a point on which I will expound in the section on criticisms of the theory. Thus, the contextual factors presented in this section are not only important in examining racial differences literature, but also is necessary in understanding the development and perpetuation of stereotype threat theory.

### Definitions of Race

Because classification is a result of subjective experience, it should come as no surprise that there is significant inconsistency in the manner by which races are categorized. Race classification has been based on language (e.g., Celtic, Slavic, Aryan) (Molnar, 1983), geography (e.g., European, African, Asian), physical characteristics (e.g., Black, White), and even religion (e.g., Jew). Unfortunately, these different approaches are often contradictory. For example, an American might assume that African and Black are synonymous. However, there are native Africans who would not fit into the American concept of "Black", such as Egyptians, Libyans, and Algerians (Frisby, 1993a). Conversely, Australian Aborigines might be perceived as "Black." How then, can we objectively define races?

One traditionally employed operational definition is based on ancestry. Its roots are in anthropology and paleontology. In this definition, each racial category is conceptualized by ancestral origin. Thus, the races are divided into three

distinct groups: Negroid (Blacks), Mongoloid (Asians), and Caucasoid (Whites). It was hypothesized that these represent three primary isolated populations in history that eventually resulted in the many contemporary races; that all current races are of these three types in whole or a mixture thereof. So, a Negroid would be defined as one whose ancestors, 40-4,400 generations removed, were born in sub-Saharan Africa. The same template can be applied to Asia for Mongoloids and Europe for Caucasoids. Thus defined, races differ in genetic material by approximately .0012%. For perspective, we share 98.5% of our genes with chimpanzees (Levin, 1997). The prevalence of an ancestral basis for racial classifications is evidenced by authors using the term African-American, which implies African ancestry, and the term Black interchangeably (e.g., Steele & Aronson, 1995).

Jensen (1998) reviewed a principal component analysis of 42 populations. The first component corresponded roughly to geographic migration distance and accounted for 27 percent of genetic variation. The second component appeared to be climate and accounted for 16 percent of the genetic variation.

Other authors have argued that race is simply a social construct and any racial classification system has no true validity (Sternberg, 2005). Regardless of how the scientific community chooses to conceptualize races, the manner in which experimental subjects are placed into racial categories for research purposes is most relevant to this study. Authors may place them in categories based on physical appearance or may rely on self-reports. According to



Herrnstein and Murray (1994), most studies in racial differences are based on how subjects choose to classify themselves.

### Racial Differences

When employing racial categories such as Black and White, which refer to skin color, it is evident that skin color may be one of the most salient differentiating factors. It is reasonable then, to scrutinize whether there are any differences between racial groups aside from skin pigmentation (color). Cognitive differences are hotly debated, as will be discussed in later sections. However, physical differences among races are frequently evidenced by literature in anatomy and medicine. These physiological differences include mean group differences in body size, body proportions, muscle/fat composition, gestation length, vital sign levels, blood type, disease prevalence, hormone levels, etc. (see Molner, 1983; Overfield, 1995; Weiss & Mann, 1981).

In addition to differences in physiology, there is also literature that presents evidence on behavioral and psychological differences. This report will summarize some that are often discussed, but it is important to remember that many of these differences are inferred from correlational data, so functional relationships are unclear.

Blacks generally are identified as having a higher prevalence of conduct/oppositional-defiant disorders and mental retardation (Levin, 1997; McDermott & Spencer, 1997). Some studies have reported these differences even after controlling for SES and other confounding factors (Herrnstein &

Murray, 1994; Levin, 1997; McDermott & Spencer, 1997). Blacks and Asians may also exhibit a higher prevalence of diagnosed schizophrenia than Whites (APA 2000, p307). The DSM IV (2000) reports that each of these pathology examples has possible genetic etiologies. However, the APA warns (2000) that such differential diagnostic trends may be influenced by clinician bias or cultural insensitivity.

There is some evidence that there may be racial differences in temperament and some personality traits. For example, there is a significant discrepancy in group tendencies toward impulsive aggression (McDermott & Spencer, 1997). There are group differences on measures of locus of control and attributional biases. For example, Blacks tend to more often agree with MMPI statements that denote externalization of blame. On self-report measures of self-esteem, Blacks tend to exhibit higher self-esteem than Whites. Also, Blacks tend to report their self-perceived academic competence more highly than whites (Levin, 1997).

On scales of neuroticism, Blacks tend to rank lowest and Asians highest with Whites in-between. On scales of extraversion, Asians tend to rank the lowest and Blacks the highest, with Whites again in-between (Levin, 1997). Levin also presented some evidence that tools for personality measure, such as the MMPI, which are often employed to effect evidence of racial differences in personality, have equal criterion validity for Blacks and Whites. Levin claims that differences in sexuality, along with aforementioned differences in extroversion

and aggression, may be related to the fact that testosterone levels are 3%-19% greater in Blacks (discrepancy range varying with age) (Levin, 1997).

This section does not offer a comprehensive summary of racial differences. Its purpose is to expose the reader to the vast body of literature that suggests a myriad of differences between racial groups. These include differences in physiology, personality, behavior, and psychology (see Herrnstein & Murray, 1994; Levin, 1997; Molner, 1983; Overfield, 1995; Weiss & Mann, 1981). The racial categories and the differences between them have had significant utility for a number of fields including history, medicine, anthropology, physiology, sociology, criminology, and psychology. Thus, one may conclude that racial groups do indeed differ in more ways than skin color.

Finally, it is important to present one caveat to these racial differences. Much of these data are correlational. It is not possible to deduce etiology from such research. Specifically, one cannot assume that such differences are either inherited or learned. The contributions of heredity and the environment likely vary across traits. For example, it is known that in the U.S., Blacks and Whites differ in prevalence rates of AIDS (Levin, 1997), sickle-cell anemia (Overfield, 1995), and diagnosed schizophrenia (APA, 2000). AIDS results from contraction of the AIDS pathogen. Thus, group difference in prevalence rates of AIDS is often believed to result from differences in environmental factors- a combination of SES, culture, education, etc. However, group difference in prevalence rates of sickle-cell anemia is known to result from differences in inherited characteristics. The etiology of schizophrenia is believed to include innate disposition and

environmental triggers. It remains unclear if group difference in schizophrenia prevalence results from innate or environmental factors. Clearly, considering these examples, neither the environment nor genes can explain the entirety of racial differences. However, uncertainty in the causes of racial differences does not invalidate their existence. Known racial differences should be recognized by school psychologists who labor to develop policies and interventions that address the over-representation of some racial groups in special education.

### Bias

Up to this point, evidence has been presented concerning the definition of racial groups and a sample of reported racial differences in current literature. Just as important as information concerning racial categorization is information about the biases and motivation of authors who argue about racial differences. The study of racial IQ differences has a long and storied history (see Gould, 1996; Valencia & Suzuki, 2001). To understand a theory like stereotype threat, one should be aware of the biases that pervade literature on racial differences. Although a complete review of the history and context of this literature is beyond the scope of this report, this section will introduce some contemporary factors that influence work in this area. The following four sections each present a factor that may be directly or indirectly related to bias in literature on racial achievement differences: Personal Bias, *The Bell Curve*, Afrocentrism, and the Pioneer Fund.

## Personal Bias

Concerning bias (in his book that addressed racial IQ differences), Gould wrote (1996): “ Science, since people must do it, is a socially embedded activity...Facts are not pure and unsullied bits of information; culture also influences what we see and how we see it” (p. 53-54).

Research on racial differences, especially racial achievements differences, is highly contentious and polarized. The argument to which an author subscribes may not be restricted to his or her objective, scientific opinion. It may also be related to that author’s identity, worldview, and self-concept (Chapman, 1993). Thus, an attack on a person’s opinion in this area may be perceived as an attack on the very way in which they perceive themselves and the world around them. Furthermore, arguments concerning racial IQ differences are related to a variety of societal and political issues (see Browne-Miller, 1995 ; Tucker, 1994). So, a person’s view on racial IQ differences may also reflect underlying political and/or social agendas.

## Bias and *The Bell Curve*

Currently, one of the most influential works on race differences in intelligence is *The Bell Curve* by Herrnstein and Murray (1994). Much of the book focuses on social and political issues related to IQ differences, but a portion presents an accumulation of research in the area of racial IQ differences. The book was not explicitly racist. It generally describes data that suggest genetic racial differences as “unfortunate” and data that suggest racial equality as

“promising.” The NY Times reviewers and many school psychologists critiqued the book quite moderately (see Braden, 1995; Frisby, 1995; Jacoby & Glauberman, 1995; Oakland, 1995). Braden (1995) points out that many trends presented in the book concerning racial differences in IQ, although unpopular, have been known to psychologists to be true for years. In December 1994, the Wall Street Journal published an editorial entitled, *Mainstream Science on Intelligence*, written by Linda Gottfredson, and signed by 52 renowned experts on intelligence to provide a consensus clarification on the current research and to address the issues discussed in *The Bell Curve*. This editorial was later published in the scientific journal, *Intelligence* (Gottfredson, 1997). The editorial delineated 25 statements which endorsed much of the contentious research findings on racial differences described in *The Bell Curve*.

However, many others were critical of the book. Editors of the NY Times, reportedly appalled by the Times’ review of the book, launched a daily campaign in the letters-to-the-editor column to destroy the book’s credibility. These editorials were often comprised of “flagrant lying about the contents of the book” (Jacoby & Glauberman, 1995, p. 331). This approach is representative of much of the response of the secular press. Those who did not lie often resorted to slander, puerile reasoning, or exaggeration. Another approach, not so overtly deceptive, comprises arguments based on ethos and emotion. A good example is Bruce McCall’s parody of *The Bell Curve*, substituting cuteness and dimples for intelligence and IQ, respectively (Jacoby & Glauberman, 1995). Michael

Novak wrote (Novak, 1994; see also Herrnstein & Murray, 1994, p 556) concerning some critics of *The Bell Curve*:

“[*The Bell Curve*’s] message cannot be true, because much more is at stake than a particular set of arguments from psychological science. A this-worldly eschatological hope is at stake. The sin attributed to Herrnstein and Murray is theological: they destroy hope.” p 59

Thus, readers should not only be aware of the possible biases of *The Bell Curve*’s authors, but the emphatic biases of those who reacted to it.

It is difficult to infer the motivation and intentions of authors who produced reactionary works that explicitly purport to discredit *The Bell Curve*, but there are many such works. For example, see Devlin, Feinberg, Resnick, and Roeder, 1997; Fischer, Hout, Jankowski, Lucas, Swindler, and Voss, 1996; Fraser, 1995; Jacoby and Glauberman, 1995; Jencks and Phillips, 1998; Kincheloe, Steinberg and Gresson III, 1996; and Neisser, 1998. Many of these books are substantial works that present informative reviews and interpretations of current data and literature. It is unclear whether any of these works had any impact on the influence of *The Bell Curve*, but they do serve as a salient reminder of the volatility of the topic. It is not the purpose of this section to provide a review of *The Bell Curve* or the reactionary works that followed. The purpose is to provide the reader with sources that show that the events surrounding the publication of *The Bell Curve* are indicative of the polarization of scientific community on this topic. When reviewing works on racial IQ differences, readers must be cognizant

of the intention of authors. Regardless of the accuracy of *The Bell Curve*, authors who report their intentions to discredit *The Bell Curve* may not present objective, balanced arguments. Indeed, contention and debate that was sparked by *The Bell Curve* prompted the American Psychological Association to appoint a task force to write an objective report (Neisser, 1996). Debate over *The Bell Curve* is an inseparable context of much contemporary literature on racial IQ differences, including literature on stereotype threat theory.

### Bias and Afrocentrism

Afrocentrists, directly and indirectly, have a significant influence on, and contribution to, literature on racial differences in IQ. Afrocentrists promote an academic agenda that focuses on the greatness of Blacks in history. Their motivation may seem innocuous but, at times, the result is a proliferation of misinformation. Clarence Walker (2001), a Black scholar of African-American studies said, “Afrocentrists have produced a therapeutic mythology designed to restore self-esteem to Black Americans by creating a past that never was” (p. xvii). Afrocentrists have included such well-known authors as Malcolm X, W.E.B Du Bois, and George G.M. James (Marable, 2000; Walker, 2001).

The roots of the problem lie in historical European presumptions about the role Africans have played in human history. More specifically, it was once a commonly discussed belief that all African achievement resulted from outside ‘Hamitic’ influence. The proposition was that non-Blacks (Egyptian, Indo-European, and Aryan) had spread across Africa and formed a small, elite ruling



class over “inferior subjects” (Howe, 1998). The decline in civilization in Africa was believed to be a result of interbreeding and deterioration of the Hematic race types. In fact, the assumption by many historical western thinkers that Africans are “inferior” is based on their belief that Africans have never created a great civilization. However, even if there is no historical evidence of a great Black civilization, such a situation does not necessitate Black inferiority. Nevertheless, African-American scholars have felt some obligation to effect evidence of Black contribution in history, and this is a primary mission for Afrocentrism. For such evidence, they turned to Egypt (Howe, 1998; Marable, 2000; Walker, 2001).

Egypt was a great civilization *and* it was located in North Africa. It was the largest, most sophisticated state up to that point in history and was probably the largest unitary state throughout its 3000 year existence. Its influence can still be seen in western culture, literature, and art. Furthermore, Egypt was the origin of such ideas as the individual soul, life after death, and a system of cosmic justice (Howe, 1998). Afrocentrists claim that Egypt was the true mother of western civilization and, most importantly, Egyptians were Black (Walker, 2001).

Afrocentrists claim that credit for intellectual developments given to Greece and Rome belongs to Egyptians. An investigation of the true contribution of Egypt lies beyond the purpose and scope of this report. However, it will be a moot point after discussing the racial origins of Egyptians.

Central to Afrocentrists’ paradigm are their claims that the Egyptians were Black. Unfortunately, space limitations restrict a more thorough explanation, suffice it to say that many paleontologists, Egyptologists, and historians agree

that Egyptians were a mixture of North Africans, European Neolithics, and Indians, although many Egyptologists find the question itself irrelevant (Howe, 1998; Walker, 2001). According to these sources, most Egyptians probably had few affinities with Blacks, or sub-Saharan Africans. However, the 25<sup>th</sup> dynasty of Egypt may have included sub-Saharan Blacks, or Nubians. Also, there have been debunked claims that Egyptians were Caucasoid (Walker, 2001).

Many of Afrocentrism's misleading claims are based on misinterpretations. For instance, they translated the Egyptian phrase, "rematch en Kemet" to mean "land of Blacks." The correct translation is "the people of the black land," referring to the black soil deposited along the Nile (Walker, 2001). It was important for Egyptians to linguistically distinguish between the Nile's "black land" (Kemet) and the desert's "red land" (deshret). Other criticized Afrocentric claims include: the 10 commandments were stolen from Egyptian ideas, Aristotle's philosophy was stolen from Egypt, Hannibal was Black, Saint Augustine was Black, Cleopatra was Black, Jesus Christ was Black, Abraham was Black, etc. (Howe, 1998; Walker, 2001). Although there have been significant contributions to literature by Afrocentrists, readers must be aware of Afrocentrism's propensity to distort facts to corroborate alternative interpretations.

Often, Afrocentric influence in scientific literature is difficult to measure without a meticulous review of the authors' citations. However, there are examples of overt employment of Afrocentric views. Janet Helm's work, which is cited in stereotype threat research (see Davis III, 2000; McKay, 1999), argues that Egyptians were Black, and because they accomplished great things, we

must reject Arthur Jensen's description of the general factor of intelligence (Helms, 1992). Many people reject Jensen's ideas, not on the basis of the evidence, but because they fear that racism may find scientific support (Miele, 2002). Such fear is native to the tenets of Afrocentrism. Jensen's work is not inherently racist (see Jensen, 1998), but advocates for African Americans, such as Afrocentrists, could perceive Jensen's work as threatening and thus act to discredit it. There may be similar resistance to other researchers who are labeled as hereditarians. When reviewing criticisms of hereditarians, it is critical to differentiate authentic, evidence-based challenges to hereditarian theory from the political struggles of groups like Afrocentrists. It is also important to gauge the influence of those politically motivated arguments on the popularity of egalitarian theories such as stereotype threat.

#### Bias and the Pioneer Fund

The work of some White scholars has also exacerbated the situation. Often criticized is the Pioneer Fund. This organization funds research on racial differences. Recipients have included Cyril Burt, Raymond Cattell, H.J. Eysenck, Arthur Jensen, J.P. Rushton, William Shockley, and Michael Levin. Herrnstein and Murray are not affiliated with the group (Jacoby & Glaberman, 1995). The work of Pioneer Fund recipients is not usually flagrantly racist or inflammatory, but often describes Blacks in a subtle, derogatory manner. This work, regardless of purpose or validity, is often not received well by Afrocentrists or equality-minded academicians. These Pioneer Fund scholars have often subjected

themselves to dangerous public anger. For instance, J.P. Rushton has to deliver his lectures by videotape because persistent threats against his life (Jacoby & Glauberman, 1995). Similarly, Michael Levin argued that there are few publications that support the hereditarian position because most editors refuse to publish such work. He reported that, because he presented hereditarian arguments, his university tried to break his tenure (the situation led to a court battle in 1991; Levin v. Harleston) and he has received death threats. He also reported that there have been movements to stifle funding for such research (Levin, 1995).

One current manifestation of this trend to censor hereditarians is the case of Chris Brand's (1996) book *The G factor* (not to be confused with Arthur Jensen's similarly titled work). The publishing house, Wiley, broke its contract with Brand by de-publishing after having published it for several weeks. The company decided that it did not want to disseminate such "repellant" views (Brand, 1996). Later, Brand was fired from his position at Edinburgh University. Despite the incendiary nature of the book, it received compliments from some academicians including Chris Chabris (1998), Hans Eysenck, Richard Lynn, Jim McKenzie, and Phillip Rushton (Brand, 1996). [note: Brand's book is only included in this report as an example of censorship]

Pioneer Fund recipients present two modes of bias of which readers should be aware. First, some recipients have been accused of publishing fabricated or misleading data. For example, some authors have accused Cyril Burt of falsifying his findings in twin studies to evidence the heritability of IQ

(Miele, 2002). Secondly, the fact that many hereditarians have difficulty funding research (outside of the Pioneer Fund) or publishing their findings means that the current body of literature could be biased in favor of egalitarians.

Pioneer fund recipients (hereditarians), like Afrocentrists, are an important component to understanding race differences. Most work on race differences (including IQ) is affected, at least indirectly, by the dichotomous struggle between egalitarians (like Afrocentrists) and hereditarians (like Pioneer Fund scholars). It is not that all of academia is engaged in a race struggle, but close investigation of citations often leads back to these factions. Also, not all Afrocentrists and Pioneer Fund scholars are motivated by biased, political ideology, but it is a contextual factor that cannot be ignored. For example, Arthur Jensen (a Pioneer Fund recipient), has been lauded for his honesty and integrity (Scarr, 1998) in a special edition of *Intelligence* (Volume 26, number 3), titled “A king among men: Arthur Jensen.” However, despite such examples, the roles of the Pioneer Fund and Afrocentrism are influenced by important contextual factors that should prompt skepticism.

### Bias Summary

Before a discussion of evidence on racial differences in IQ, which is central to stereotype threat research, it is necessary to review the contextual factors presented thus far. First, inconsistency in the definition of race was discussed, but it was shown that there are widely accepted racial differences in phenotypic characteristics of physiology and possibly in behavior or certain

personality variables. Second, it was argued that some authors' scientific beliefs are intimately related to their perception of themselves and their worldviews. Discussions about topics such as race differences may incite anxiety or bias in such people. Third, publication of *The Bell Curve* has polarized some of the academic community. Several contemporary works have been published either to support or refute Herrnstein and Murray's hereditarian position. Fourth, it was explained that some White scholars and some Black Afrocentrists may be motivated to propagate misinformation to promote a personal agenda that diminishes our ability to objectively examine the issue.

Superficially, this section on contextual factors may seem tangential and irrelevant. However, an honest examination of a theory concerning racial differences in IQ without the inclusion of these contextual factors is analogous to discussing evolution in the early 1900s while denying the influence of theology. In fact, within this topic, the motives and biases of writers may be just as important as their data.

## Cognitive Ability

### IQ

IQ scores are the primary diagnostic criteria for mental retardation, and the tendency of Blacks to perform below average on these tests contributes substantially to their over-representation in special education. Stereotype threat theory attempts to explain why Blacks may perform lower on such tests.

An individual's IQ is a standard score in the statistical sense of the word standard. IQ is not a direct measure of a person's intelligence, but it is an estimate of an individual's general intellectual functioning. It is derived by comparing the individual's performance to the performance of people of the same age on various cognitive tasks. Historically, IQs are reported as standard scores that have a mean of 100 and a standard deviation of 15-16. For perspective, two standard deviations above the mean is a score of 130, which is the lower threshold for "genius" level functioning. Two standard deviations below the mean is a score of 70 and is the upper threshold for mental retardation (Sattler, 2001). Approximately 95% of the population have scores within 2 standard deviations of the mean (between 70-130) (Neisser et al., 1996). One standard deviation below the mean is a score of 85. This is the alleged mean IQ of Blacks.

Contrary to frequent criticism, IQ tests exhibit fairly strong reliability and validity. IQ scores are statistically reliable for children as young as four (Mash & Barkley, 1996). They also exhibit predictive validity and are the best predictors of academic achievement (Jensen, 1998; Oakland, 1995; Sattler, 2001). The correlation between IQ and school grades is about .50 (Neisser et al., 1996). IQ is correlated with a variety of advantageous life outcomes (Herrnstein & Murray, 1994; Oakland, 1995). For example, intelligence scores are correlated with job performance. The correlation estimates range from .30 to .50 (Neisser et al., 1996). Also, intelligence scores are negatively correlated with juvenile crime (about  $r = -.19$ ) (Neisser et al., 1996).

IQ is quite stable throughout one's lifetime. Statistics have shown that IQ measures a cognitive ability that is relatively resistant to change. Scores obtained at age 12 and at age 18 are correlated at about .89 (Neisser et al., 1996). Furthermore, each year of formal education adds, on average, only one IQ point to the expected adult IQ (for comparison, standard errors of measurement are often as high as 10 IQ points) (Sattler, 2001).

Compensatory programs, such as Head Start, have attempted to improve the prognosis for disadvantaged youths. These programs often have some short-term success but improvements on tests usually fade within two years (Neisser et al., 1996; Sattler, 2001). This is because, although these children are learning more information and skills, the rate at which they learn is only artificially increased. After the program ends, the rate returns to its natural level. It is this "rate of learning" that is more associated with IQ than accumulated knowledge (Oakland, 1995).

For this report, it is important to know that 40%-80% of variance in IQ has been reported to be due to inherited factors (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; Braden, 1995; Herrnstein & Murray, 1994; Levin, 1997; Oakland, 2001). IQ is comparably valid for persons of different social and racial groups (Oakland, 1995). Furthermore, the general factor (g factor) accounts for at least 50% of the variance in IQ (Jensen, 1998).

Finally, it is important to note the fact that, for reasons that are not clear, mean IQ scores for populations tend to rise over time and have done so in every industrialized nation. This phenomenon is known as the Flynn Effect after James



Flynn, who has published a numerous works addressing it (Jensen, 1998; Neisser, 1998). The increase in average IQ scores over time manifests most significantly on highly g-loaded tests. The greatest rise has been found in data from Raven's Matrices, a non-verbal, highly g-loaded test. For tests such as the Raven's, IQ scores rise as much as 20 points per generation (30 years) (Jensen, 1998; Neisser, 1998). Scores have risen about equally for both Blacks and Whites over the past sixty years, while the two groups have maintained their average one standard deviation difference. Thus, Blacks in the 1980's performed at the level of Whites in the 1930's (Neisser, 1998). In contrast to the Flynn Effect, scores on achievement tests have declined over time (Jensen, 1998; Neisser, 1998).

### G factor

Individuals differ in their performance on different measures of intelligence. A person who excels at non-verbal matrices will not necessarily excel at verbal items. However, tests that measure one type of ability tend to be positively correlated with tests of other cognitive abilities. For example, a person who excels at matrices will, more often than not, perform above average on verbal IQ items. Psychologists use factor analysis to clarify such relationships. Factor analyses of IQ tests reveal a general factor of intelligence (the g factor) that accounts for variance across all types of IQ tests and subtests. Specific factors that are common to specific types of tests are also revealed (Jensen, 1998). Theorists differ in whether their focus emphasizes the g factor or specific

factors. One common view incorporates both types hierarchically with the g factor at the apex. There are significant discrepancies in theorists' interpretation of the g factor. It has been described as a simple statistical regularity, a mental energy, abstract reasoning ability, or neural processing speed. There are theorists that argue that the g factor is misinterpreted. They prefer to examine profiles of strengths and weaknesses based on group factors (Neisser et al., 1996). In Snyderman and Rothman's (1987) survey of over 600 experts, 58% favored some form of general intelligence whereas only 13% of the experts favored separate faculties. Sixteen percent thought the data was too ambiguous to judge.

The g factor can be difficult to define because it is an abstract, statistically derived property of IQ tests. Factor analyses of tests do not always produce general factors like they do for IQ. For example, factor analyses have not shown a g-factor for personality measures. For measures of IQ, however, the g factor is almost always present. Some items on IQ tests (and some entire tests) are more correlated with the g factor than others. Such correlations are referred to as "loadings." This is an important concept for arguments about racial differences because racial group differences in performance vary with differential g loadings. G-loaded items generally call for the education of relations and correlates, they require deductive or inductive reasoning, and make minimal demand for specialized or esoteric knowledge (Jensen, 1998). There are several current theories that describe additional cognitive abilities that are more specific than the g-factor. These factors account for variance in IQ performance after variance due to the g-factor has been partialled out. Researchers have used such results

of factorial analysis to produce hierarchical models of intelligence with the g-factor at the apex. Gf-Gc theory (also referred to as CHC theory), for example, describes two strata of abilities under the g-factor. For an extended discussion, see McGrew and Flanagan (1998).

### Black/White IQ Gap

It is widely believed that, regardless of the cause, there is approximately one standard deviation difference in mean performance between Blacks and Whites on IQ tests and other measures of cognitive ability (Frisby, 1995; Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). More specifically, Blacks exhibit a mean IQ of 85. This is not a new idea. Sir Francis Galton, in his grossly inappropriately titled chapter (in *Hereditary Genius*), *The Comparative Worth of Different Races*, estimated a 20.9 IQ point deficit for Africans (Jensen, 1998). According to a few studies, the Black/White gap may be narrowing by about 0.2 standard deviations. Also, some studies have reduced the Black/White difference by about 1/3 by controlling for SES (Herrnstein & Murray, 1994). It has, however, been demonstrated that the traditional 1 SD gap may be present in children by about three years old even though IQs are not even considered reliable until five years of age (Levin, 1997). Native Africans tend to exhibit a mean IQ 70 or, 2 SD below the White average, on various types of cognitive ability tests (Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). This trend, in the US and in other countries, is most prominent on nonverbal, g-loaded tests. Not all minorities exhibit depressed performance on IQ tests. East Asians typically have an IQ that

is 0.5 SD above the White mean of 100 (Herrnstein & Murray, 1994; Jacoby & Glauberman, 1995). The Asian/White difference is most pronounced in tests of visiospatial ability, not verbal ability (Herrnstein & Murray, 1994).

Because western IQ tests are designed by members of western culture and because normative data are often based on a representative White majority, one may question whether differences are due to cultural bias in test items. One answer to this question has been termed, the “Spearman hypothesis.” Charles Spearman, the discoverer of the g factor, remarked (in *The Abilities of Man*) that the Black/White difference was most marked in those tests known to be saturated with g (Jensen, 1998). This hypothesis was termed Spearman’s Hypothesis and was tested thoroughly. Studies have confirmed that the Black/White difference is wider on items that appear to be culturally neutral and is wider on items most correlated with the G factor (see Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997) This explicitly contradicts the theory that race differences are a result of culturally biased IQ tests. An example of this trend is differential performance on the digit span subtest of the Weschler intelligence tests. In the first section of the test, the subject must repeat a sequence of numbers. In the second section of the test, the subject must repeat a sequence of numbers in reverse order (backwards). Both tests have minimal cultural loading, they only require a basic ability to pronounce numbers. The second portion of the test, the backwards portion, is twice as g loaded as the forward section. It requires more “g” to simultaneously remember and mentally reverse the numbers. The Black/White difference is about twice as great on the

backward section compared to the forward section (Herrnstein & Murray, 1994). Both sections are administered consecutively to every subject, so this is also strong anecdotal evidence that motivation is not a factor (since it is unlikely that, when starting the backwards test, Black subjects become spontaneously unmotivated).

Raven's Matrices is an IQ test employing abstract figures and are generally thought to have little correlation with cultural factors and is heavily loaded on g. The Raven's test consistently produces a 1 SD Black/White difference in American subjects and a 2 SD deficit in native Africans (Herrnstein & Murray, 1994; Levin, 1997). Conversely, tests that appear to be heavily culturally biased actually serve to reduce the Black/White difference and often overpredict Black performance on certain criteria. Many of those tests require retention and recall of esoteric knowledge (Herrnstein & Murray, 1994; Noble, 2003).

Some psychologists, such as George M. Harrington, have suggested that the approach to item selection for IQ tests is flawed. He hypothesized that, because items are chosen based on the performance of a group that is usually comprised mostly of Whites, that items are biased against Blacks. This hypothesis has been contradicted by experimentation. An example was a study performed using the K-ABC. This test is widely employed to test Blacks because there is an abnormally small Black/White difference in IQ scores for the K-ABC (Mayfield & Reynolds, 1997). In congruence with Spearman's Hypothesis, the K-ABC has been shown to be a much less valid measure of the g factor than other

IQ tests (Jensen, 1984). In the study of Harrington's Hypothesis (invalid item selection), two forms of the K-ABC were developed, a White form and a Black form. Each form was developed to maximize the performance of the respective ethnic group. Contrary to Harrington's Hypothesis, Blacks scored lower than Whites on both forms. This result has been replicated using a number of different IQ tests (Hickman & Reynolds, 1986; Mayfield & Reynolds, 1997).

There still remain two, less inferential possibilities for showing cultural biases of IQ tests against Black Americans. If one could remove the effects of low SES status from Black samples or find a White sample that is affected by factors similar to low SES, one might find the necessary evidence. Both of these possibilities have been investigated to some extent.

Many studies have turned to African populations in hope of removing confounds such as oppression and cultural deprivation that are alleged to be present in African American populations. Often there have been attempts to control confounding factors by sampling Africans that are employed, live in urban areas, have finished school, and by using culturally neutral tests such as Raven's Matrices. Even then, African IQ usually exhibits a mean of 1.5 SD to 2 SD below the White average (Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). If the results of such studies had shown a mean of 100, it would have strongly supported claims that IQ tests are biased against oppressed African Americans.

The second possibility mentioned previously was examining White children who are deprived of opportunity, as are many Blacks. There has been a unique study on such a population. Deaf children suffer from extreme

deprivation. They suffer from auditory deprivation, profound delays in language exposure, inconsistent and nonstandard language models, greater prevalence of abnormal genetic conditions, and frequently dysfunctional family dynamics. Showing that White, deaf children exhibit low IQs would provide excellent evidence that culture and environment should be a prime suspects in the Black/White IQ gap. Unfortunately, this is not the case. White, deaf children exhibit a mean IQ of 99.97 (Braden, 1995).

Studies have shown evidence that the racial IQ discrepancy is not attributable to formal test characteristics such as verbal/nonverbal; individual/group; culture-loaded/ culture-reduced; or race of examiner. Furthermore, it is unlikely that the difference is due to esoteric knowledge or testing familiarity. Additionally, a black/white racial IQ difference has manifested in every country for which representative data are available (Jensen, 1998). There is excellent factorial similarity between Black and White standardization and IQ is comparably valid for persons of different racial groups (Jensen, 1998; Oakland, 1995) Finally, Blacks with mixed ancestry typically exhibit less depression in IQ than Blacks without mixed ancestry (Jensen, 1998).

It should be noted that the trends in underperformance of Blacks covered in this section represent trends in central tendency. As such, there are evident exceptions to the norm. It has been shown that there are Blacks who achieve above expected academic levels across the range of SES levels. Also, it has been reported that, as Blacks' SES improves, the achievement gap narrows (Robinson-Heath, 2002).

To summarize, a preponderance of evidence shows that the depression of Blacks' IQ scores is due to factors other than systematic bias in IQ tests. The reasons presented include: The IQ gap is exacerbated by culturally neutral and g-loaded items and tests; IQ tests do not under-predict and may over-predict Black performance on several criteria (academic achievement, college success, etc.); Blacks perform more poorly than Whites on test items that were chosen to maximize Blacks' performance; there is excellent factorial similarity between Black and White standardization samples; the discrepancy exists across many cultures/countries; test characteristics (other than g loadings) do not predict the discrepancy (see Herrnstein & Murray, 1994; Jensen, 1998; Kush et al., 2001; Levin, 1997).

### Introduction to Contemporary Explanations

A variety of explanations have been forwarded to explain discrepancy in average performance between racial groups on criteria that are related to cognitive abilities. In this report, theories are organized according to their assumption about the primary origin of the discrepancy. Three categories of origins are noted.

The first possible source of variation comprises genetics or inheritance. This position holds that much of the performance variation between groups is due to factors that are innate, or inherited. Such explanations will be referred to as Hereditarian Theories, and the proponents of such theories are often called Hereditarians.



The second possible source of variation comprises environmental factors that effect actual differences in ability between racial groups. This position holds that individuals from different racial groups tend to be exposed to different types of environments, but within the racial groups, there is some consistency in environment. Characteristics of the environment are assumed to account for the largest amount of variance between groups. Such explanations will be referred to as Environmental Deficit Theories, and the proponents of such theories are often called Egalitarians.

A third type of possible source of variation comprises factors that vary between racial groups and affect the validity of the measurement instruments. This position holds that there are non-genetic differences between racial groups and, because of these factors, the instruments used to measure racial differences are biased. Such explanations will be referred to as Cultural Relativism Theories and the proponents of such theories are also called Egalitarians.

It is important to note that those who do not support the hereditarian perspective are collectively referred to as Egalitarians. Egalitarians espouse that, at least within the cognitive domain, racial groups are inherently equal. Among egalitarians, however, there is an important distinction. Those who endorse Environmental Deficit Theories believe that, due to environmental factors, African Americans have not developed cognitively to their full potential. Thus, these Egalitarians, like Hereditarians, believe that the measured differences are real. Conversely, Egalitarians who subscribe to Cultural Relativism Theories believe

that the measured differences are not real but are artifacts of biased or invalid measures.

All reasonable psychologists believe that intelligence results from an interaction between inheritance and the environment (Gould, 1996). Hereditarians do not deny the influence of environmental factors and Egalitarians do not deny the influence of genes. Thus, the controversy and the difference between contemporary theories relates to differing beliefs in the relative contributions of genetics and the environment to racial differences. Furthermore, none of these categories of theories are necessarily mutually exclusive. They may each describe factors that directly contribute to the problem and/or interact with other factors to contribute.

In Snyderman and Rothman's survey on expert opinions (1987), 94% of experts agreed that there is evidence for significant heritability in IQ. According to the survey, experts, on average, believed that 60% of the variation in IQ is due to genetic factors. This result corroborates what is often purported: that 40% to 80% of variance in IQ is due to genetic differences (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; Braden, 1995; Herrnstein & Murray, 1994; Levin, 1997; Oakland, 2001). Also in Snyderman and Rothman's (1987) survey, 45% believe that the difference in performance between White and Black Americans on IQ tests is due to both genetic and environmental variables. 15% believe it is entirely environmental and 1% believes it is entirely genetic. The following sections present samples of each of the three aforementioned categories of theories. The samples provided are not exhaustive and only serve as examples.

## Hereditarian Theories

The hereditarian perspective has pervaded psychological science since its inception (see Herrnstein & Murray, 1994, Gould, 1996). The once celebrated Sir Francis Galton, who laid foundations for differential psychology, empirical psychology, statistics, and eugenics, popularized the tenets of the hereditarian approach in his 1896 publication, *Hereditary Genius* (Gould, 1996; Herrnstein & Murray, 1994; Jensen, 1998; Viney & King, 2003).

Hereditarians believe that a significant portion of the racial IQ gap may be due to innate differences between racial groups. Examples of contemporary hereditarian theorists include: Jensen (1998), Herrnstein and Murray (1994); Levin (1997); and Rushton (2003). Some Hereditarians believe that the three primary races may be placed on a continuum across which scores of human characteristics vary. For example, some purport that Mongoloids (Asians) may tend to exhibit higher IQs than Caucasoids (Whites), and Caucasoids in turn, exhibit higher IQs than Negroids (Blacks). It is argued that many other traits vary consistently across this racial continuum, although not all traits lie in the same direction. Such traits include rate of maturation, brain size, athletic ability, sex hormones, temperament, and twinning rate. Furthermore, Out-of-Africa models of human origins have been proposed to explain such patterns (see Levin, 1997; Rushton in Lieberman 2001).

Perhaps one of the most extreme examples of hereditarian ideology is the hypothesis posited by Pioneer Fund recipient J.P. Rushton. Rushton argues that

the IQ difference between Whites, Blacks, and Asians is the result of a complex interaction between genetic properties of: brain size, genital size, rate of sexual maturation, length of menstrual cycle, hormone levels, fertility, altruism, etc. His position is precariously based on evolution, biological notions of reproductive strategies, and parental investment. A simplistic paraphrase would be: compared to Blacks, White children, evolutionarily speaking, must be smarter because Whites are predisposed to have far fewer children and invest significantly more parental resources (Herrnstein & Murray, 1994). It is unclear whether there is any empirical support for such a theory.

According to the APA task force (Neisser et al., 1996), certain aspects of brain anatomy and physiology may be related to intelligence. These aspects include arborization of cortical neurons, cerebral glucose metabolism, evoked potentials, nerve conduction velocity, sex hormones, and others (Neisser et al., 1996).

Regardless of whether many hereditarians endorse such radical ideas like Rushton's, they all believe that genetics plays a fundamental role in IQ differences. Much of the evidence to support claims of genetic influence is based on twin studies (e.g., Bouchard et al., 1990). Researchers compare identical twins (monozygotic) reared together, identical twins reared apart, fraternal (dizygotic) twins reared together, and fraternal twins reared apart. Furthermore, unrelated children reared together or apart are sometimes studied. Comparing identical twins, which share 100% of their inherited traits; fraternal twins, which share 50% of their inherited traits; and unrelated children, researchers are able to

study the influence of genes. Similarly, by comparing these types of children when they are reared together with others who are reared apart, researchers are able to control for differences in environmental factors. Recent twin and adoption studies suggest that variance due to environmental differences between families is substantial in early childhood, but it becomes quite small by late adolescence (Neisser et al., 1996). For some critics, however, some adoptive environments are not different enough to provide credible evidence.

Heritability estimates increase with the age of the subjects. For children, heritability estimates are as low as .45, but by late adolescence heritability estimates are around .75. However, this heritability does not imply immutability. Height, which is a trait that has high heritability, continues to rise on average with successive generations. Similarly, mean IQ scores show an increasing trend over time (Neisser et al., 1996).

One important caveat of the hereditarian perspective is the difference between within-group and between-group heritability. Some hereditarians are criticized for overextending the applicability of some heritability estimates to explain between-group variance. Although a trait, like IQ, may have a high rate of heritability, such a condition is not sufficient to infer that differences between groups (races) are inherited. For example, the height of corn has high heritability. If one were to plant a field of corn and then observe that some plants grew very short, it would be reasonable to assume that genetics played a role in the variance observed. However, if one were to plant two fields of corn, one in fertilized soil and one in rocky, dry soil, the plants in the rocky field are likely to

grow shorter despite their inherited disposition to grow to the same height as those plants in the fertilized field. This would happen even if the plants in both fields were genetically identical strains. Thus, there would be environmental factors that accounted for a substantial portion of between-group variance (rocky vs. fertilized). Yet, these environmental factors would have minimal influence on the initial planting of a single field (within-group variance). For more exhaustive coverage of this topic, see Herrnstein and Murray (1994), Jensen (1998), and Neisser (1998).

### Environmental Deficit Theories

Theorists who endorse Environmental Deficit Theories believe that individuals from different racial groups tend to be exposed to different types of environments. They believe that developing in impoverished environments may have deleterious effects on the cognitive development of many African Americans.

Craig Frisby (1995) lists several possible oppressive environmental influences on Blacks that may result in Blacks' depressed cognitive ability and that are often discussed in literature: legacy of slavery, teacher prejudice, inadequate schools, lack of parental involvement, poverty, lack of learning opportunity, deficient mother-child interaction, and lack of academic role models. Additionally, he lists several often used explanations relating to psychological maladjustment: lack of motivation, learned helplessness, low self-esteem, and

negative peer pressure. Each of these environmental factors could have an effect on Blacks' cognitive development.

Environmental Deficit Theories that emphasize the impact of ecological variables on racial differences in cognitive development may be congruent with some well-known developmental theories. These include the work of Jean Piaget, Albert Bandura, Lev Vygotsky, and Urie Bronfenbrenner (see Bandura, 1993; Bronfenbrenner, 1993; Goldhaber, 2000).

Of particular interest to this report is the work of Bandura (1993) whose work defines constructs that might serve as alternatives to stereotype threat. For example, Bandura describes the role of self-efficacy on cognitive development and academic performance. According to Bandura's expectancy-value theory (Bandura, 1997), greater expectancy that certain behavior can secure specific outcomes and greater value placed on those outcomes lead to increased motivation to perform the activity. This belief that one's behavior can secure desired outcomes is related to one's perceived self-efficacy. Bandura suggests that people's self-efficacy has a global affect on their functioning. This includes how they think, motivate themselves, feel, and behave. Individuals with high self-efficacy, compared to other individuals of the same ability but less self-efficacy, set higher aspirations, show greater strategic flexibility, and more accurately evaluate their own performance. Furthermore, individuals with high self-efficacy, compared to other individuals of the same ability, are more persistent, efficient, and are less likely to reject correct solutions prematurely. Bandura contends that efficacy beliefs independently contribute to intellectual performance and are not a

simple reflection of such performance (Bandura, 1997). There are many parallels between the trends described by Bandura's research and those described in stereotype threat research. It is possible that stereotype threat manipulations (discussed later) interact with perceived self-efficacy.

It could be reasonably argued that Bandura's theory is more similar to Cultural Relativism Theories. However, it is listed in this report as an Environmental Deficit Theory because Bandura's ideas describe relatively fixed attributes in individuals that would affect their performance across a variety of domains, not just IQ testing. Thus, the implication that some environmental variable set predisposes some individuals to such cognitions and performance is qualitatively different than Cultural Relativism Theories which imply that the performance measures (IQ tests) are not valid.

Another, less articulated theory is suggested by the research of James Flynn and others on the Flynn Effect. It is known that Blacks exhibit IQ gains over time at about the same rate as Whites so that Blacks today may perform, on average, as well as Whites did a generation ago (see Neisser, 1998). It is suggested by Flynn, Neisser, and others that this is evidence against hereditarian theories. Attempts to uncover the factors that explain the Flynn Effect have defied many intuitive hypotheses. According to Flynn, examination of research shows that the effect cannot be substantially accounted for by any of the following: increased test sophistication, altered test-taking strategies (the Brand Hypothesis), changes in nutrition, SES, urbanization, disease eradication (part of the Storfer Hypothesis), historical trauma (e.g., WW II), improved



preschool, TV, and improvements in education. Some theorists speculate that by uncovering the causes of the Flynn Effect, we may understand the factors that lead to differential ability between races to perform on IQ tests. These theorists maintain that, according to evidence surrounding the Flynn Effect, these factors must not be genetic (see Neisser, 1998; Jensen, 1998).

James Flynn has posited that one type of environmental influence that could lead to individual differences in performance is what he calls multipliers (Flynn, 2003). He suggests that relatively small differences in inherited ability can develop over the lifetime into substantial differences via multipliers. As an example, he describes that someone with a slightly better than average athletic ability may be predisposed to like sports, practice more, be recruited for team play, and otherwise engage in activities that result in significantly greater than average ability over time. He suggests that there may be analogous multipliers for IQ. He believes that, with just a slight advantage in initial cognitive ability, a person could be predisposed to environments that magnify the difference. Such a person might be more inclined to enjoy school, receive praise, read, be admitted to advanced classes, etc. This theory clearly illustrates how, in congruence with other Environmental Deficit Theories, factors in the environment can facilitate development of real differences in cognitive ability.

## Cultural Relativism Theories

Those who subscribe to Cultural Relativism Theories believe that there are sociocultural factors that vary between races such that the racial IQ gap is actually an artifact of biased measurements.

Craig Frisby (1995) listed several possible environmental factors often discussed in literature that may invalidate traditional cognitive assessment of Blacks. One such factor is cultural bias in tests. This possibility may stem from a lack of Blacks in standardization samples, Blacks' preference for dynamic vs. static testing, item loading on white middle class culture, different race of examiner, or lack of "test-wiseness". Another category of Frisby's suggestions relates to a discrepancy between Afrocentric homes and Eurocentric schools. This category includes opposition to White cultural values, teachers' cultural incompetence, deficits in multicultural curricula, ebonics language preference, and different behavioral/learning styles.

Expectancy theory describes Black underperformance as a result of lowered expectations (Clark, 1955; Ogbu, 1986; also summarized in Brown, 2001). This theory argues that Blacks internalize negative stereotypes about Black intellectual inferiority. This process then causes a depression in their performance expectations, which in turn negatively impacts actual performance. This is in contrast with Stereotype Theory, which does not require internalization of stereotypes or lowered expectations about one's ability to perform.

Robinson-Heath (2002) found that, among low SES Blacks, those who were classified as high achievers tended to exhibit higher academic motivation

and higher success expectancy. These variables were measured by self-report questionnaires. She argued that high motivation and expectations result in more positive outcomes for these subjects. It is unclear whether academic success effected higher expectations and motivation, or vice versa. However, the study does present the possibility that some Blacks may not be performing optimally due to depressed motivation/expectations. Robinson-Heath did not include Whites in the study, so it is unknown whether these variables affect races differentially.

For their part, some Afrocentrists maintain that the Black/White IQ gap is a result of IQ test bias against Black Cultural Learning Styles (BCLS). However, a number of studies, including a 1987 meta-analysis of over 7,000 studies (see Frisby, 1993a), have shown that cognitive/learning styles of any kind are not robust explanations for differential educational outcomes. Learning Styles (including BCLS) generally lack construct validity, criterion validity, predictive utility, and there is little reliability in instruments used to measure such things (Frisby, 1993a). Furthermore, BCLS bias in testing rests on the assumption that White scholars (test writers), because they are White, cannot understand Black culture, but Black scholars (Afrocentrists) understand all cultures. Some scholars have even suggested that the manner in which Afrocentrists deny their westernization is in itself characteristically western (Frisby, 1993b).

The System of Multicultural Pluralistic Assessment (SOMPA) was developed by Jane Mercer to provide less discriminatory methods to be used by practitioners in assessment of minorities. Specifically, she wanted to minimize

the over-reliance on standardized IQ for decision making with minority children. Mercer argued that it would be inappropriate to judge human behavior via a singular statistical distribution because of diverse lifestyles, cultures, personalities, etc. Similarly, because of the diffuse effects of culture, we should not evaluate cognitive performance by a single distribution. Thus, SOMPA assumes that there are distinct normal curves for different ethnicities and that minorities should be evaluated by different distributions. This pluralistic ideology holds that, because IQ tests produce different means for different ethnicities, they are *ipso facto* biased. Mercer used Estimated Learning Potential (ELP) to modify minority students' WISC (Weschler's IQ test) scores. This technique reduces the numbers of minorities that qualify for certain diagnoses, like mental retardation, by substituting an Estimated Learning Potential for a standard IQ score. A number of experts doubted the validity and integrity of such a process, most notably Joan Goodman (1977). In the following years, several studies were conducted to investigate ELP's validity (e.g., Beck, 1984; Wurtz, Sewell, & Manni, 1985) and were generally critical of the technique. Jirsa (1983) said:

The statistical manipulation of current performance (WISC-R IQ) may therefore succeed in eliminating certain children from special programming, but that in no sense changes the child in terms of his or her current functioning. The ELP process is descriptive, not prescriptive- it does not provide any strategies, by itself, for increasing a child's school related competency.(p 19)

In the past decade, SOMPA's prevalence in literature has sharply dropped. Nevertheless, its residual influence on arguments about racial IQ differences remains relevant.

John Ogbu's cultural ecological theory has received considerable attention and was summarized in the APA task force report on intelligence (Neisser et al., 1996). His theory is a response to hereditarian and environmental deficit theories (Fordham & Ogbu, 1986, p. 180). John Ogbu argues that caste-like minority status negatively impacts academic achievement. Caste-like minority status is differentiated by Ogbu from autonomous minorities who are not economically or politically oppressed. He also distinguishes caste-like status from voluntary minorities, or immigrants, who perceive themselves to be in an improved situation or expect their situation to improve (Ogbu, 1978; Ogbu, 1994). According to Ogbu, Blacks perceive a restriction on their academic resources and social/economic mobility due to factors beyond their control. They may perceive some of these factors as being controlled by the White majority. According to Ogbu, this situation leads to a number of debilitating variables: distrust of the school system, frustration with restrictions, adoption of survival strategies that are incongruent with skills needed for academic achievement, low self-esteem, etc. (Fordham & Ogbu, 1986).

Academic achievement may be perceived by blacks as a White domain. Pursuit of such academic achievement may be derogatorily perceived by blacks as "acting white". Such efforts may be interpreted as a betrayal of Black solidarity or racial identity. This belief leads Blacks, as a caste-like minority, to oppose academic striving, both socially and psychologically (Fordham & Ogbu, 1986). This culture of anti-intellectualism may explain some of the variance in racial differences between caste-status Blacks and White majorities.

## Stereotype Threat

In 1995, Claude M. Steele and Joshua Aronson published the seminal article, *Stereotype threat and intellectual test performance of African Americans*. Although theorists had often discussed the possible impact of prejudice and stereotype (see Allport, 1954; Goffman, 1963), Steele's and Aronson's stereotype threat theory was unique in that it explained the behavior via immediate situational threat and did not assume internalization of stereotypes or images. This theory fits most appropriately in the Cultural Relativism camp. Steele and Aronson supported their hypothesis with results from a series of five experiments, all of which were presented in the 1995 article. The phenomenon they demonstrated would later be replicated by dozens of experimenters. Based on Psychinfo searches (in January 2007), over 235 works have been published concerning stereotype threat since its 1995 inception. Of those sources, only a few authors, such as Arthur Whaley and Paul Sackett, were explicitly critical of the theory, (and Sackett's colleagues) (see Sackett, Hardison, & Cullen, 2004a, 2004b; Sackett, Schmitt, Ellingson, Kabin, 2001; Whaley, 1998). Some authors were so enamored by the theory, that they have begun instructing educators on how to apply it (see Aronson, 2002; Aronson, Fried, & Good, 2002; Good, Aronson, & Inzlicht, 2003; Steele, Spencer, & Aronson, 2002).

The 1995 publication came on the heels of the controversial *The Bell Curve* publication (Herrnstein & Murray, 1994). Egalitarians, Afrocentrists, the secular press, and many academicians who feared a resurgence in racism were

clamoring to effect evidence of environmental explanations for racial differences in IQ (see section on *The Bell Curve*). For those who were searching for reasonable alternatives to hereditarian ideas, stereotype threat theory may have seemed like a powerful and elegant solution. Not only does it describe an environmental influence that differentially affects races' academic performance, it does so without assuming any actual difference in cognitive ability.

The community's appreciation for stereotype threat theory was evidenced when Claude Steele was bestowed with the APA's 2003 Distinguished Scientific Contribution Award and the APA's 2003 Distinguished Contributions to Psychology in the Public Interest Award for his work on the topic (Kersting, 2003). Stereotype threat theory was popular enough to be covered by the secular media. It was discussed in an edition of the television program, *Frontline* (Chandler, 1999; see also Sackett, Hardison, & Cullen, 2004), in the television program *20/20* (ABC network, September 15, 2006) in the motion picture, *The Perfect Score* (Robbins, 2004), and in the bestselling book *Blink* (Gladwell, 2005). Dr. Steele has given guest lectures on the theory and has even provided expert legal testimony concerning the theory's implications (Expert, n.d.). Dr. Steele's stereotype threat research (and the research of many other ST theorists) has been financially supported by numerous grants. Grants listed on his curriculum vitae which were likely used to primarily investigate stereotype threat and its implications sum to \$1,484,425 (Steele, n.d.).

The original authors offer the following definition of stereotype threat:

“When a negative stereotype about a group that one is part of becomes relevant, usually as an interpretation of one's behavior

or an experience one is having, stereotype threat is the resulting sense that one can then be judged or treated in terms of the stereotype or that one might do something that would inadvertently confirm it.” (Steele, Spencer, & Aronson, 2002, p 389)

In addition to this definition, a few other important characteristics were delineated. First, stereotype threat is a situational threat which means that the elicitation of stereotype threat depends on several variables. The negative stereotype must be made relevant and salient through environmental cues. More specifically, the task by which the subject will be evaluated must be related (by the subject’s perception) to the negative stereotype. Furthermore, that relationship must be salient to the subject. Also, it is not necessary for the subject to believe the stereotype, only for the subject to believe that the stereotype exists and may be believed by individuals who could evaluate the subject’s performance. The authors also state that the more a subject identifies with the stereotyped group (i.e., Blacks), the more he/she should experience stereotype threat. They also believe that the more a subject identifies with the domain of performance (e.g., academics, intelligence), the more he/she should experience stereotype threat (Steele, Spencer, & Aronson, 2002; see also Steele, 1997; Steele & Aronson, 1995).

Generally, it is theorized that when an African American takes an IQ test, he/she may become cognizant of negative stereotypes about Black intellectual inferiority. Research has shown that these stereotypes remain widely known (for example, see Devine & Elliot, 1995). Stereotype threat infers that the subject, now aware of the stereotype and its relevance to a task which he/she is about to perform, becomes concerned about confirming the stereotype. It is suggested



that this fear may be related to the possibility of confirming the stereotype to oneself or permitting others to confirm it through one's actions, thus facilitating the stereotype's perpetuation. This fear of stereotype confirmation allegedly depresses the subject's performance.

Researchers use specific elicitors (or manipulations) in experimental settings to induce stereotype threat. There are two types of these elicitors that are commonly used. One type of elicitor is test description. Specifically, the test is described as diagnostic of a person's ability. This is referred to as the Diagnosticity manipulation. The second type of elicitor is called the Race Priming manipulation. This elicitor involves asking the subject for a self-report of race prior to the test (see Steele & Aronson, 1995). Each of these elicitors has been shown to significantly depress the performance of African Americans in experimental samples (Steele & Aronson, 1995; Claytie Davis III, 2000, Joseph Brown, 2001).

Theorists speculate that this depression in performance is mediated by one or a combination of the following: distraction, narrowed attention, anxiety, self-consciousness, withdrawal of effort, over-effort, lowered expectancies and/or dejection emotions (Keller & Dauenheimer, 2003; Stangor, Carr, & Kiang, 1998; Smith, 2004; Steele, 1997; Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). Schmader and Johns (2003) showed evidence that stereotype threat manipulations may reduce working memory capacity. Other studies have shown that stereotype threat manipulations may increase blood pressure of African

Americans (Blascovich, Spencer, Quinn, & Steele, 2001) and stereotype threat may be mediated by testosterone levels (Josephs, Newman, & Brown, 2003).

Smith (2004) reviewed the research on the behavioral mediators of effort and self-handicapping and concluded that there was insufficient empirical support for these variables as complete or partial mediators. Smith (2004) also reviewed phenomenological mechanisms of anxiety, evaluation apprehension, performance confidence, stereotype endorsement, perceptions of the test, and feelings about self. Smith reported that support for anxiety and performance confidence as partial mediators emerged in some studies. Generally, Smith reported that research on mediating mechanisms is inconclusive and that, at this point, the possibility of these mechanisms should not be prematurely dismissed. Smith's review of mediating mechanisms was not confined to studies on racial achievement differences but also included research on gender differences and research on athletic tasks. Although often conceptualized as a single phenomenon, this author located no research to explicitly indicate that gender stereotype threat and racial stereotype threat function via the same mechanisms.

There are a number of variables that are known to increase the strength of the stereotype threat phenomenon. In 2002, Steele, Spencer, and Aronson drew on a number of different studies to summarize what was known about these variables. First, higher domain identification is known to increase stereotype threat. For the purposes of this report, that means that Blacks who view cognitive ability as important to their identity will be more vulnerable to stereotype threat. The second variable of interest is group identification. More specifically,

Blacks who perceive their race as an important characteristic of their identity, who feel some sense of solidarity with other Blacks, or who view themselves as being part of that group may be more susceptible to stereotype threat. The authors (Steele, Spencer, & Aronson, 2002) caution that conflicting results have been detected in studies that investigate this variable. Thirdly, task difficulty influences stereotype threat. The stereotype threat phenomenon has been shown to be greatest on tasks that are more difficult. Fourth, the apparent diagnosticity of the task has been shown to increase stereotype threat. When Blacks view a test as diagnostic of their ability (usually because it was described as such), they tend to exhibit more stereotype threat. Fifth, stereotype relevance increases stereotype threat. If Blacks are told that a test is biased against their race, they should tend to exhibit more stereotype threat. The sixth factor was not listed explicitly by the authors, but is derived from the original study (Steele & Aronson, 1995). This factor is stereotype salience. For example, having to report one's race prior to the task increased the stereotype threat phenomenon for Blacks. Presumably, this is because it raised their awareness of the possible stereotype. Additionally, it has been demonstrated that stereotype threat is not unique to Blacks or to cognitive tasks. Stereotype threat has been demonstrated in white males with implied comparison to Asians (Aronson, Lustina, & Good, 1999; Smith & White, 2002), in Latinos (Gonzales Blanton & Williams, 2002), and in athletic tasks (Stone, 2002; Stone, Lynch, & Sjomeling, 1999). Also, it has been shown by Italian (Cadinu, Maass, & Frigerio, 2003), German (Keller & Dauenheimer, 2003), and French (Desert, Croizet, & Leyens, 2002) researchers.

The effects of stereotype threat manipulations on women, which usually involve elicitation of the stereotype of female inferiority in mathematics, are extensively documented: Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Smith & White, 2002; Spencer, Steele, & Quinn, 1999; Stangor, Carr, & Kiang, 1998; Walsh, Hickey, & Duffy, 1999.

In the first two studies of their 1995 report, Steele and Aronson used an ANCOVA design. The first variable was race, of which there were two levels, Black and White. The second variable was diagnosticity. This refers to the description of the test provided to the subjects in the instructional set. Some subjects were given a diagnostic description by which they were led to believe the test was diagnostic of their verbal ability. Others were provided a nondiagnostic test description in which they were not informed that the test was diagnostic of their verbal ability. In Study 1, a third level of diagnosticity was used in which the task was framed as nondiagnostic, but the subjects were challenged to take the task seriously. This nondiagnostic-challenge condition was dropped in the second study. The dependent variable (the task) was performance on a series of verbal GRE items. This performance was measured by the number of correct responses. The covariate used during statistical analysis was each subject's prior performance on the SAT.

According to the authors, stereotype threat theory predicts a significant interaction effect. Specifically, the diagnostic condition (DC) should significantly depress Blacks' performance relative to their performance in the nondiagnostic condition (NDC) or nondiagnostic-challenge condition (NDCC). This diagnosticity

manipulation should not affect White's performance. Study 1 failed to yield a conventionally significant interaction ( $p < .08$ , excluding NDCC), but Study 2 did yield a significant interaction ( $p < .05$ ).

Manipulating the test description was not the only way that the authors depressed Black performance. In Study 4, all subjects received the NDC. One group of these subjects was given a questionnaire prior to the test that concluded with an item on which they could record their race. This is known as the race-prime group. The other group was given the same questionnaire prior to the test, except that the final item (race) was omitted. This was the non race-prime condition. Again, the SAT covariate was used.

This simple, one-item race-priming manipulation produced a significant interaction effect ( $p < .01$ ). Racial-primed Blacks performed worse than non racial-primed Blacks. Although not predicted, Whites may have exhibited depressed performance in the non-racial prime condition. The authors failed to report whether the effect of priming on Whites was significant.

Claytie Davis III (2000), a student of Aronson, replicated part of Steele's and Aronson's (1995) studies but used a Black experimenter and did not use a White comparison group. The study reported a significant effect of diagnosticity ( $p = .033$ ) and a significant effect of racial priming ( $p = .037$ ). The mixed results of the study did not support Davis' hypothesis that racial identity attitudes can predict academic performance of African Americans in the stereotype threat condition.

Joseph Brown (2001), a student of Steele's, replicated Steele's and Aronson's 1995 study and found a significant race x diagnosticity interaction ( $p < .01$ ). Also, Brown concluded that expectancy does not mediate stereotype threat because in his sample, there were no significant racial differences in expectancy, but the stereotype threat effect was still replicated. Expectancy was measured by one 7 point Likert item preceding the test. The validity and reliability of such a method is unclear.

Patrick McKay conducted an experiment to test the effects of stereotype threat on Blacks' performance on Raven's Matrices, which are highly g loaded. McKay found insignificant results ( $p = .42$ ) on the race by diagnosticity interaction (McKay, 1999). McKay was able to produce conventional levels of significance by covarying using SES estimates such as income and by manipulating his sample by excluding subjects whose responses on questionnaires suggested that they were not convinced of his diagnosticity manipulation. Both procedures generate analyses that are ungeneralizable. Furthermore, his use of a covariate is inappropriate because there is a significant between-group difference in SES. Misusing covariates will be discussed further in proceeding sections. McKay also used the dataset from his 1999 dissertation in some subsequent publications.

### Criticisms of Stereotype Threat

Stereotype threat theory presents a robust, replicated laboratory phenomenon. However, the interpretations of these findings are problematic.

This section will present some of the problems with stereotype threat theory as an explanation for depressed academic performance of Blacks.

The design of many stereotype threat studies, including Steele and Aronson's 1995 study, employs ANCOVA. In Steele and Aronson's study, the authors used ANCOVA to remove variance due to racial differences in performance on standardized tests of verbal ability (SAT scores). This equalized the groups statistically to provide a more accurate measure of differences in performance due to the stereotype threat manipulations (diagnosticity, racial-priming, etc.). In the studies reviewed by this author, Whites, in congruence with nationally established trends (Herrnstein & Murray, 1994), exhibited higher scores on the covariate (e.g., Brown, 2001; Steele & Aronson, 1995). McKay (1999) used parental income as a measure of SES for the covariate instead of prior performance on a task similar to the DV. Nevertheless, Whites exhibited higher scores on that covariate as well.

There is no inherent problem with this design (ANCOVA) as a tool to determine whether the manipulated variables are responsible for variability in the DV. The problem lies in its prevalent, fallacious misinterpretation. Steele and Aronson present their findings in a manner that is easily misunderstood. They lead readers to believe that the nondiagnosticity manipulation equalized group performance. For example, Aronson states (2002, p283):

Moreover, there was no difference between the performance of the black test takers under no stereotype threat and that of white test takers.

Joseph Brown (2001) interprets findings similarly when he stated that the nondiagnostic condition eliminated the performance gap (p. 4). Brown further states, "...while the [non-diagnostic condition] led to equal performance between Whites and African Americans" (Brown, 2001, p22). However, this equalization was done artificially through statistical techniques. All variance related to pre-existing race differences in the groups' verbal performance was removed by the covariate, in which Whites outperformed Blacks. When comparing races, insignificant difference in means of the authors' ANCOVA design should be interpreted as indicating that the differences between the groups were not significantly different, except for the pre-existing race differences (which were removed). Thus, the design and results have no direct applicability to the problem or nature of Blacks' academic/achievement/IQ deficit.

Furthermore, in interpreting the results of the ANCOVA design, Blacks in the DC performed worse than Whites in the DC after covarying. So, in addition to pre-existing race differences (which were removed), the racial gap was further widened. The requisite conclusion is that the stereotype threat manipulations (elicitors) may anecdotally depress performance for some Blacks in a manner similar to this laboratory study, but that the induced deficit in Black performance exists independently of pre-existing deficits. The implication is that removing stereotype threat manipulations (e.g., test diagnosticity, racial priming) in real-world settings will not remove racial differences in IQ/achievement or equalize races' performance. However, purposeful inclusion of such manipulations could be used to further exacerbate the problem and widen the gap. Thus, the



stereotype threat phenomenon is real, as evidenced by the empirical studies, but has questionable social utility.

Another method (which leads to the same conclusion as above) of conceptualizing the ANCOVA design, as applied to stereotype threat, is to recognize that the tests for significance were applied to residual scores. These are scores from which pre-experiment (pre-existing) individual differences were statistically removed. Thus, if two group means are found to be significantly different, then such difference has manifested independently of the pre-existing differences.

This interpretation of the results, that stereotype threat increases the racial gap, is dependent on all of the statistical assumptions of the ANCOVA design being met. However, it is possible that the assumptions were violated. Steele and Aronson (1995) did not publish sufficient data to examine this issue. Several possible violations were summarized by Wicherts (2005). First, domain identification and ability may be positively correlated, resulting in an interaction between the covariate and the Diagnosticity manipulation. This would create a curvilinear relationship and regression weights that vary across cells. Second, if stereotype threat functions through mediators, such mediators could also violate homogeneity of the regression weights. Third, it is assumed that the covariate is error free, which is not necessarily the case. Also, stereotype threat could have affected performance on the covariate. Therefore, the precision of the analysis could have been compromised (Wicherts, 2005).

The correct interpretation of the means in the stereotype threat analysis may have been more apparent to readers if the authors included unadjusted means for each of their performance-related studies. However, no such data are provided. It is likely that Whites outperformed Blacks in all conditions, but this is masked by statistical techniques (covariance). Whether or not the NDC truly equalized group performance would be more accessible to readers of this literature if unadjusted means were reported. Instead, the authors have chosen to present their data in such a way that necessitates a meticulous statistical scrutiny to understand the exact nature of the results. The failure of many academicians to exhibit such skepticism has resulted in a widespread misunderstanding of the stereotype threat phenomenon.

Recently, Paul Sackett, Chaitra Hardison, and Michael Cullen published an article addressing this issue of widespread misinterpretation (Sackett et al., 2004a). The article reported the following: 90.9% of scientific articles sampled characterized the stereotype threat incorrectly; 87.5% of popular media articles characterized the stereotype threat incorrectly; 56% of applicable introductory psychology texts sampled characterized the stereotype threat incorrectly. In response, Steele and Aronson (2004) denied that there was cause for concern, but as Sackett et al (2004b) point out, Steele and Aronson do agree that the alleged false interpretations are indeed incorrect.

An analogous problematic interpretation is evident when Brown (2001), a student of Claude Steele, attempted to examine the role of expectancies on stereotype threat, he measured subjects' perceived self-efficacy via a Likert style

self report item after the subjects complete a verbal task. He purports to find that Blacks perceived themselves as having higher verbal skills than Whites perceived of themselves. However, this claim is based on means that controlled for SAT and pre-study self reports. Brown stated that there was no statistical difference in pre-study self reports. However, Whites' mean SAT score was significantly higher than Blacks'. Thus, in an absolute sense, Blacks may not have reported higher verbal skills. In fact, it is impossible, outside the confines of statistical analysis of this singular study, to generalize this particular data to any population. This finding concerning a main effect of perceived efficacy was not the focus of Brown's work (although the rest of the study did use covariance in an analogous way), but it typifies the vulnerability of his work to misinterpretation. It should be noted that some studies (see Cokley, 2002) have shown that Blacks may tend to report higher levels of self-esteem than Whites, which may be related to Brown's self-efficacy. Nevertheless, the data provided by Brown to his readers is insufficient to permit the interpretation he renders.

A point that may further confound interpretations of stereotype threat studies is the as-of-yet unexplained and irregular phenomenon where the stereotype threat manipulation removal (i.e., NDC, no race-prime) depresses White performance. For example, see Study 2 and Study 4 (Steele & Aronson, 1995), Hypothesis 2 (McKay et al., 2002), Study 1 and Study 2 (Brown 2001). Often the significance of this trend and its relative contribution to the interaction effect are not discussed. This trend may be related to the fact that subjects in the NDC were explicitly told that similar tests are biased in favor of Whites, but the

test being administered has been developed to remove this bias (Brown, 2001). The implied message to White subjects is that they may not do as well on the presented test as they may do on other tests. It was not discussed how this affected White performance, but the available data suggests that their performance was affected. Furthermore, because this trend increases the significance of the stereotype threat interaction effect, it may facilitate the overestimation of the stereotype threat effect's influence on Black subjects.

Inconsistent findings of conventional statistical significance is another problem concerning stereotype threat phenomena. Stereotype threat theory predicts a significant interaction of race and stereotype threat manipulation in ANOVA or ANCOVA designs. Certainly, several studies have evidenced significant results, but there are some that have not (by conventional levels). For example, see Study 1 (Steele & Aronson, 1995) or Hypothesis 2 (McKay et al., 2002).

An inherent problem of stereotype threat theory is the failure of the theorists to incorporate, or sometimes even acknowledge, the known trends that characterize the depressed achievement of Blacks, of which the theory purports to explain a substantial proportion. These trends include the Spearman hypothesis, over-prediction, discrepancies beginning in toddlerhood, pervasiveness of the gap in other cultures, etc. (see section on Black and White IQ differences).

Another facet of the racial gap that confounds stereotype threat theory is the trend of East Asians to out-perform Whites (Herrnstein & Murray, 1994;

Jacoby & Glauberman, 1995). The stereotype threat phenomenon has indeed been measured in White subjects by suggesting to the White subjects that their performance will be compared to Asians (Aronson, Lustina, & Good, 1999; Smith & White, 2002). These studies illustrate the broad applicability to the theory. No one, however, has tried to interpret this stereotype threat laboratory finding in the manner that findings which involve African Americans have been interpreted. If these findings (White/Asian) were applied in the inferential manner of others (White/Black; Male/Female), one would conclude that White American children, on average, underperform relative to East Asians on some cognitive tasks because they (White American children) are preoccupied with confirming racial stereotypes. Discussing the Asian/White differences may tend to invoke less egalitarian sensibility than discussing Black/White differences. Thus, the weakness of the theory's utility should be more readily apparent. It is unreasonable to assume that the cognitive performance of White children in America is incessantly undermined by awareness of a stereotype against them. It is not even clear if White school children are aware that such a stereotype exists. Similarly, research has shown that the performance of White athletes may be undermined by via stereotype threat manipulations (Stone, 2002; Stone, Lynch, & Sjomeling, 1999). Stereotype threat manipulations have been shown to affect practice effort (Stone, 2002) and actual performance (Stone, Lynch, & Sjomeling, 1999). As of yet, no one has seriously postulated that the persistent trends of Whites underperforming or being underrepresented in some sports (Levin, 1997; Stone, Lynch, & Sjomeling, 1999) is due to their fear of confirming stereotypes of

Black athletic superiority. There is certainly no evidence that such prevalent and pervasive fear exists in the minds of White athletes across the US. These examples, Asian/White IQ stereotype effects and Black/White Athletic stereotype effects, illustrate the weakness in overextending stereotype threat findings in the manner that theorists have done for the Black/White IQ gap.

There has been some attempt to test the applicability of stereotype threat theory to real world settings in which Blacks tend to underperform on cognitive tasks. One such study was conducted by Cullen, Hardinson, and Sackett (2004). The authors used previous findings to develop a couple of models with which to test stereotype threat theory. One model incorporated the suggestions of Steele and others that the phenomenon manifests predominantly among the individuals in the upper end of the distribution. Another model additionally addressed the possibility that those individuals at the extreme upper tail of the distribution may not be as affected because the task was not challenging enough. The authors used regression to test whether their large data sets showed trends predicted by the models. No support for stereotype threat was found. However, despite the large sample and meticulous statistical analyses, the study was not an experimental one. Also, the authors noted that some components of the study might be interpreted by others as allowing for stereotype threat phenomenon. However, the authors contend that when the entirety of their results are considered, there is substantial evidence to show that no stereotype threat phenomena were present, or at least not in the manner traditionally theorized.

Finally, the most important caveat of stereotype threat research that should elicit skepticism from school psychologists is the population in which the stereotype threat phenomenon manifests. Experimental studies examining stereotype threat have usually relied on college-aged subjects (see Aronson, Lustina, & Good, 1999; see also Brown, 2001; Davis III, 2000; McKay, 1999; Steele & Aronson, 1995). As discussed earlier several variables are thought to contribute to the effect of the stereotype threat phenomenon. These include domain identification, group identification, knowledge of the stereotype, and stereotype salience. Steele points out (1997) that it is high achieving students that should be most affected. Thus, it is reasonable to question whether stereotype threat theory applies to the African-American students who underperform on IQ tests and contribute to over-representation in special education; a problem facing many school psychologists. These African-American students are not the requisite high achievers and it is not likely that they exhibit the domain identification described by Steele. For very young, school-aged children, it is unclear whether they would meet the other criteria as well, such as group identification and knowledge of the stereotype. Also, for some underperforming African-American students, they may comprise a majority in their particular school. So it is also unclear how salient the stereotype would be in testing situations in those facilities.

### Subsequent Research on Stereotype Threat

As discussed earlier, a number of researchers have conducted experiments similar to Steele and Aronson's 1995 study. Many of these do not significantly deviate from the original 1995 methodology and do not circumvent the inappropriate use of a covariate. There are, however, two notable exceptions. First, Loveless (2000) conducted a large sample study examining the effect of stereotype threat on the ACT exam. This study was commissioned, supervised, and published by the research division of the organization that publishes the ACT. Second, Brown and Day (2006) conducted a study to specifically address the criticism of covarying on prior performance in stereotype threat research.

With the support of the ACT organization, Loveless investigated whether racial priming influences Blacks' and Whites' performance on the ACT differently. 4 levels of priming were employed in the experiment: (1) prime 2 weeks before and immediately before the test, (2) prime immediately before the test, (3) prime 2 weeks before the test, and (4) no priming. All other variables were held constant, including the simulation of an actual ACT test and all of the implicit influences of that context. Subjects were categorized according to race (Black and White) and community type (urban, suburban, and rural). Data from 1,030 subjects were analyzed, providing for a powerful analysis. Loveless found (1) no statistically significant differences among priming levels, (2) no statistically significant interactions between priming levels and race, and (3) no statistically significant interactions between priming levels and community type. There were statistically significant differences for race and community type. While Loveless



examined these findings separately for math and reading scores, the aforementioned pattern of findings was identical across both domains. While racial priming has been shown to have a significant effect in traditional stereotype threat research, this study strongly indicates that, in a real world context and in the absence of erroneous interpretation, racial priming (a primary stereotype threat elicitor) has no impact on the substantial variability in performance between racial groups.

Brown and Day (2006) examined the stereotype threat effect on performance on the Raven's Advanced Progressive Matrices (APM). Unlike many previous researchers, Brown and Day did not covary out between-group variability so that their results could be generalized to the racial gap. Their sample was 53 Black participants and 83 White Participants from a Midwestern university. The authors note that the university is less selective than the university sites of much previous research, and consequently, their results may be more generalizable to the general population. Raven's Advanced Progressive Matrices (APM) is highly g loaded IQ test, and traditional research has shown a mean score of 100 for Whites and 85 for Blacks (Jensen, 1998). Brown's and Day's data were collected within a 2 (race) x 3 (instructions) between-groups design. Levels of race were Black and White and were measured by self report. Levels of instructions were high threat, standard threat, and low threat. During the data collection procedure for the high threat condition, experimenters referred to the APM as an IQ test and indicated that it measures participants' ability. During the data collection procedure for the standard threat

condition, the standardized instructional language designated in the test manual was used. During the data collection procedure for the low threat condition, test items were referred to as puzzles, and any suggestion that the test was evaluative was removed. Brown and Day also collected self-report ACT scores.

The authors found a significant interaction effect indicating the manifestation of the stereotype threat phenomenon. They also found a significant main effect of race. The authors conclude that their study “offers strong support for the hypothesis that race differences in cognitive ability test scores could be accounted for with a simple, contextual variable that is independent of biological factors and even test content” (Brown & Day, 2006). The authors’ implicit perseveration on a Stereotype Threat – Hereditary dichotomy is a salient reminder of the issues of egalitarian bias discussed earlier.

While the authors did not use ACT scores as a covariate, they did employ an unusual analysis

“Specifically, we first conducted simple regression analyses within each race, regressing APM scores on ACT scores across experimental conditions and saving the residuals for further analysis. Second, after adding the race-specific mean performance levels to the unstandardized residuals of these separate regression analyses, we submitted the residualized test scores to a 2 (race: African American or White) x 3 (instructions: standard threat, high threat, or low threat) between-groups general linear model analysis of variance (GLM-ANOVA).” (Brown & Day, 2006; p 981)

This procedure reduces within-group (error) variance, while isolating between group variability and not adjusting the group means for performance on previous tests, like the ACT. Considering the large body of literature that shows a robust, significant, pervasive racial gap on cognitive ability measures (Frisby, 1995;

Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997), any study that relies on statistical manipulation to show significant findings and then claims to account for that racial gap should be scrutinized. Furthermore, the assumption that stereotype threat elicitors will only affect mean between-group differences and not differentially affect within-group variability is unfounded. Because stereotype threat depends on personal characteristics like domain identification, it would be appropriate to speculate that, in addition to affecting mean performance, it increases within-group variance for Blacks by strongly influencing some of Black participants, but not others. Statistically removing this variance may mask the nature of stereotype threat, produces results that are ungeneralizable to previous research on racial differences, and implies that the manipulations cannot be used to equalize group performance in a real world setting.

While the authors' (Brown and Day 2006) analysis may have been inappropriate considering their conclusions, the unrepresentative sample that they obtained renders criticisms of their analysis moot. Previous authors have used a covariate to equalize performance of racial groups a priori. This poor design (ANCOVA) was heavily criticized, as discussed earlier. Brown and Day suggested that they avoided this issue and their results are generalizable to the racial gap. However, whether purposeful or coincidental, Brown and Day, for all practical purposes, equalized groups a priori via unrepresentative subject selection.

The reported mean IQ for Blacks is 85, while the mean for Whites is 100 (Frisby, 1995; Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). This

significant discrepancy (15 points) defines the racial gap and it is what stereotype threat authors purport to account for with their experimental manipulations (see Brown, 2001; Aronson 2002; Brown & Day, 2006). The data reported by Brown and Day were analyzed by this author to examine whether this racial gap was present in their sample. The Raven Manual (Raven, Raven, & Court, 1998b) was referenced for all conversions. Raw scores reported by Brown and Day were converted to percentiles using Table APM21 (US Smoothed Un-timed Norms), the 18-22 age range, and conventional rounding rules. The percentiles were then converted to IQs using Table APM36. When a percentile rank fell between two IQ scores, the higher score is reported. The results are reported in Table 1.

Table 1  
*Raw Score to IQ Score Conversion for Brown and Day 2006 Article*

<u>Test Description</u>	<u>White</u>			<u>Black</u>		
	Raw	%ile	IQ	Raw	%ile	IQ
Low threat	22.44	57	103	24.29	64	106
Standard threat	24.28	64	106	22.42	57	103
High threat	24.67	67	107	19.41	46	99

Under no experimental condition did Brown and Day detect a racial gap that was representative of what is typically measured. All means across all conditions were in the average range (90-110). The greatest discrepancy (8 points) was observed in the high threat condition. The discrepancy in the standard threat condition, which is generalizeable to racial gap research because it uses standardized language, was only 3 points. Even if Brown's and Day's manipulations account for 100% of the between-group variance, which is

improbable, then these manipulations would only account for a 3 point racial gap. Clearly, their data provide no support for the hypothesis that the typical 15 point racial IQ gap which manifests in the general population “could be accounted for with a simple, contextual variable.” Further, these IQ conversions were based on un-timed norms. If timed norms were available, the scores would be elevated and appear even less representative of the population to which the authors would like to generalize their results. As with all prior stereotype threat studies, the unrepresentative sample and statistically manipulated data analyzed by Brown and Day (2006) provide no support for stereotype threat as an explanation for the depressed performance by Blacks on cognitive performance tests outside of laboratory settings.

## CHAPTER III

### METHODS

To investigate the differential effect of the stereotype threat phenomenon on cognitive performance across racial groups, a two-factor, between subjects, 2x3 analysis of variance (ANOVA) design was planned. Independent variables consisted of race (Black and White) and task description (stereotype threat, no stereotype threat, control). The stereotype threat group was analogous to the DC group used in previous research, the no stereotype threat was analogous to the NDC group used in previous research, and the control group was unique to this study. The control group simulated a typical standardized testing situation. The dependent variable was performance on the Cognitive Task measured by accuracy. ACT scores were collected for use as a covariate to compare the results of this study's analysis of variance design to the results from an analysis of covariance design, which is often used in stereotype threat research. A post-test questionnaire was used to collect demographic data and contained a multiple choice item that served as a deception check.

## Participants

80 participants were recruited from Oklahoma State University (OSU) and 80 participants were recruited from Langston University. Students were offered extra credit for participation. At OSU, undergraduate students were recruited from courses in the Educational Psychology department. At Langston, undergraduate students were recruited courses in the Psychology department. Participating students were assigned to the experimental condition (stereotype threat, no stereotype threat, control) as a group by course section. Group administration occurred during each course sections' regular meeting time. Students were provided with a commensurate alternative to participation.

Students at Langston primarily comprised African-Americans. Students from Langston who self-reported race as African-American or Black were assigned to the Black condition. Students from Langston who did not self-report race as African-American or Black were excluded from analysis but were permitted to participate. Students at OSU primarily comprised Caucasians. Students from OSU who self-report race as Caucasian or White were assigned to the White condition. Students from OSU who do not self-report race as Caucasian or White were excluded from analysis but were permitted to participate.

## Procedure

### Overview

In each experimental condition, participants were asked to: (1) Listen to the experimenter explain the purpose of the study and the content of the consent forms; (2) sign a consent form to participate in the study; (3) listen to a task description; (4) complete identification information on the response sheet; (5) listen to instructions for the verbal section; (6) complete items on the verbal section for 15 minutes; (7) listen to instructions for the nonverbal section; (8) complete items on the nonverbal section for 15 minutes; (9) sign a consent form to permit the acquisition of the participant's ACT score from the registrar's office; (10) complete a post-test questionnaire; and (11) receive a flyer announcing a future debriefing session and providing experimenter contact information. The entire procedure was group administered by experimental condition to students in a university course.

### Detailed Procedure

Prior to the experiment, approval was obtained from the institutional review board of (IRB) Oklahoma State University (OSU) and Langston University (LU) (see Appendices D and E). Also, letters were obtained from the registrar's office at both universities indicating cooperation with the acquisition of participants' ACT scores.



Three course instructors of undergraduate developmental psychology courses at OSU agreed to permit recruitment from their respective class sections and to provide credit for participation. These three OSU course sections were randomly assigned to the three task description conditions while ensuring that each of the task description conditions was assigned to one course section.

Three course instructors of undergraduate psychology courses at LU agreed to permit recruitment from their respective class sections and to provide credit for participation. These three LU course sections were randomly assigned to the three task description conditions while ensuring that each of the task description conditions was assigned to one course section.

Experimenter teams recruited participants and performed the experiment procedure during the regularly scheduled meeting times of university courses. The experimenter teams included three to four experimenters. At least one experimenter was White and at least one experimenter was Black in each condition. The primary investigator, who delivered the verbal instructions and task descriptions was a white male. The other experimenters assisted by distributing forms, proctoring, and collecting forms.

Once the class session began, the course instructor informed the students that they could participate in the study for extra credit, but their participation was voluntary. A commensurate alternative was provided for student who wanted credit but did not want to participate. No student in any condition elected to participate in the alternative.

Participants were then given a packet which contained (1) the study consent form, (2) the cognitive task, (3) the response sheet. Characteristics of each of these forms varied across levels of task description (TD). Participants were not permitted to view a particular form until directed to do so. All forms given to a particular participant contained a unique identifier code to ensure that responses were attributed to the correct participant. Participants were not aware of the identifier code.

Next, the primary investigator verbally explained the contents of the consent form. These contents varied across TD levels to facilitate the TD deception. Participants were then asked to read and sign the consent form. Experimenters were available to answer questions and any participant who chose to not participate would be allowed to leave. Additionally, any participant who decided to discontinue participation at any time during the rest of the experiment would be permitted to return to class. No participant discontinued participation at any time during the study.

After each participant signed the consent form and the forms were collected, the primary experimenter orally delivered a description of the Cognitive Task. To facilitate the TD deception, the task description was dependent on the condition to which the participants' course was assigned.

Then, the participants were asked to complete identification information items on the response sheet. To facilitate the TD deception, participants were prompted to provide different information depending on the TD level. Next, the participants were given instructions for completing

the verbal cognitive task section, were informed of the time limit (15 minutes), and then worked on that section. After the time limit expired, instructions were provided for the nonverbal cognitive task section, the time limit for that section was disclosed (15 minutes), and then the participants worked on the nonverbal section. Experimenters proctored the task and ensured that participants were working on the appropriate section, were not communicating with each other, and were not visibly cheating. After the time limit expired, experimenters collected the cognitive task and response sheets.

Next, the experimenters distributed the registrar consent form and the post-test questionnaire. These forms were identical across all experimental conditions. The primary investigator verbally explained the contents of the consent form, which requested access to participants' ACT scores from the registrar's office. Participants were then asked to read and sign the consent form. Experimenters were available to answer questions, and any participant who chose to not provide consent would be allowed to continue with the study and would receive credit. All participants consented. Then, the participants completed the post-test questionnaire. After all of the participants completed the questionnaire, they were informed that they would be debriefed at a later time, were given the debriefing flyer, and the experiment was concluded.

## Task Description Manipulations

The Task Description variable comprised three levels: stereotype threat, no stereotype threat, and control. Because this study is exploratory and because some statistical power would be lost with the omission of the covariate, a constellation of manipulations were employed to facilitate the deception instead of a single elicitor. It was most important to this study that the stereotype threat phenomenon was elicited, and it was not designed to isolate the effect of individual stereotype threat elicitors.

The stereotype threat condition (STC) used two elicitors from previous research known to cause the phenomenon to manifest: (1) the task was described as diagnostic of participant ability and (2) participants were asked to report their race (called racial priming) on the response sheet prior to the cognitive task. The no stereotype threat condition (NSTC) was analogous to the non-diagnostic condition or challenge condition in previous research. Participants were challenged to do their best, but the task was never described as diagnostic and the term ability or similar terms were not used. Also, the participants were not prompted to report their race prior to the cognitive task. The control condition (CX) was designed to simulate a typical standardized testing situation. The stereotype threat elicitors (diagnosticity and priming) were not presented in the CX. Participants were told that the task was similar to the ACT and GRE, and that their scores could be used for school admittance, scholarships, and employment in the future.

A constellation of manipulations were employed to support the deception of task description across each of the three levels. These differences included differences in: (1) the title of the study indicated on the study consent form, the cognitive task, and the response sheet; (2) the name of the cognitive task indicated in verbal descriptions, on the task protocol, and response sheet protocols; (3) the covers of the cognitive task protocol (see appendices A, B, and C) (4) the running head on pages of the cognitive task; (5) the format of the response sheet (see appendices H, I, J, and K); and (6) the data requested on the response sheet (see appendices H, I, J, and K).

For the STC, the name of the study was “Understanding Personal Factors that Affect Verbal Ability and Intelligence,” the title of the task was “Diagnostic Test of Verbal Ability and Intelligence,” and the running head on the task was “VERBAL ABILITY TEST” for the first section and “INTELLIGENCE TEST” for the second section. For the NSTC, the name of the study was “Understanding Different Methods for Solving Problems,” the title of the task was “Problem Solving Challenges,” and the running head on the task was “PROBLEM SOLVING CHALLENGES.” For the CX, the name of the study was “ISOT Compensatory Provision for Oklahoma Partners and Group Data Collection,” the title of the task was “Integrated Scholastic-Occupational Test,” and the running head on the task was “ISOT. © 2006 ETS . Institutional Version 3C.”

For the control condition, a response sheet was constructed that appeared to be a scantron, or computer scored protocol. Participants were required to “bubble-in” responses. For the STC and NSTC, the participants wrote the letter that corresponded to their response choice on a blank line. Also, for the STC, there was an item that prompted subjects to record their race.

The language in the task description for the control was designed to convince participants that the task was an actual standardized test with legitimate contingencies (e.g., school admittance, scholarships, employment). The verbal task description that was delivered for the control condition was:

“We have a unique opportunity for you in lieu of our usual research work. [Site name] is a development partner for ETS, the company that publishes the SAT and GRE exams. ETS, along with collegiate partners have developed a new test called the ISOT. ISOT stands for the Integrated Scholastic-Occupational Test. ISOT scores may be used like SAT, ACT, or GRE scores. These uses include university admittance, graduate school admittance, scholarships, etc. Additionally we project that it may be used by examinees when seeking employment in the private sector, and certainly for city, state, and federal jobs. This test has the added benefit of being very short when compared to traditional assessments.

In appreciation of [site’s name]’s work in the ISOT’s development, they have agreed to provide the test free of charge to 500 [site’s name] students across the next 5 years. This is also research because we will be conducting some group statistical analyses. The university and ETS are not interested in your individual scores and we will not provide individual students’ scores to any person or institution as a part of this research.

After the test, we can provide you with your scores. If you would like to have your scores considered official, you can do so by written request or online request. If you want your scores to be official and to be sent to a location in the future for school

admittance, scholarship, or employment, you can file an online request. There may a charge per request to send scores out.”

You should do your best work in order to help us in our analyses and so that these scores might help you in future academic and occupational opportunities.”

The language used in the stereotype threat condition is similar to that used by Steele and Aronson (1995) and Brown (2001) but had to be adapted to the instrument used in this procedure. The language was designed to elicit the stereotype threat phenomenon as it manifested in pervious research. The verbal task description that was delivered for the stereotype threat condition was:

“The purpose of this study is to assess your verbal reasoning skills and fluid intelligence. Also, we are very interested in personal factors that affect verbal and fluid intelligence abilities. So we will ask you to respond to some questions about yourself before and after the exam.

The test is quite challenging in order to provide a genuine test of each subject's strengths and limitations. Because of the test's difficulty, you may expect to not get many of the items correct. The test must be difficult to provide a genuine evaluation of your personal verbal ability, intelligence, and limitations so that we might better understand the factors involved in such.

After the test, we can provide you with feedback that may be helpful to you by familiarizing you with some of your strengths and weaknesses. You should do your best work in order to help us in our analysis of your verbal and intellectual ability.”

The language used in the no stereotype threat condition is similar to that used by Steele and Aronson (1995) and Brown (2001) but had to be adapted to the instrument used in this procedure. The language was designed to avoid eliciting the stereotype threat phenomenon as it manifested in pervious research, but to otherwise approximate the language in the STC as closely as possible.

The verbal task description that was delivered for the no stereotype threat condition was:

“The purpose of this study is to assess psychological factors involved in solving verbal and abstract problems. We will also look at whether the mode of administration (that is paper or computer) makes a difference.

The test is quite challenging because our research is focused on how people solve difficult verbal problems. Because of the test’s difficulty, you may expect to not get many of the items correct.

After the test, you will be given feedback to familiarize you with the kinds of problems that appear on tests you may encounter in the future. You should do your best work in order to help us in our analysis of the problem solving process.”

#### Instrument

The Cognitive Task comprised two sections. The first section was composed of 32 items from the verbal section of the Graduate Record Examination (GRE). Most items were identical to those used by Brown (2001) and similar to items used by Aronson and Steele in their 1995 study (Brown, 2001). Brown’s sample passage and passage discussing slavery were omitted, and another passage from a GRE practice test was added. Also, Brown’s three anagrams were omitted and a section of five antonym items were added. The antonym items were taken from a GRE practice test.

Steele and Aronson (1995) permitted participants to work on these verbal GRE items for 30 minutes, while Brown (2001) permitted participants to work on the items for 25 minutes. However, the scheduled times for the university courses from which participants were recruited for this study were limited, and a



15 minute time limit was necessary for the verbal section. To compensate for the abbreviated time limit, the order of some of the items was adjusted. Specifically, participants worked on the reading passage items last because reading passage items require more time per item than the other types of items on the section. The task required participants to complete items in the following order: (1) 8 items requiring participants to select a word or pair of words that best fits blanks in a given sentence; (2) 5 analogy items; (3) 5 antonym items; and (4) 14 reading comprehension items relating to 4 reading passages. All items on the verbal section were multiple choice items with four response choices.

The second section of the cognitive task comprised 30 items from the Raven's Advanced Progressive Matrices (APM) Set 2. APM is an intelligence test originally developed by J.C. Raven based largely on Spearman's conceptualization of the general factor of intelligence. The test was designed to assess educative mental ability and to minimize demand on verbal or cultural knowledge. As described by Raven, Raven, and Court (1998), studies have consistently shown APM to be one of the best measures of the general factor of intelligence. The retest reliability of APM for adults is high ( $r = 0.91$ ) (Raven, Raven, & Court 1998b). The retest reliability coefficient ( $r$ ) is 0.86 for 11.5 year olds, is 0.76 for 10.5 year olds, and is 0.86 for Black American Students. Split-half reliability coefficients vary between 0.83 and 0.87. However these reasonable split-half coefficients were obtained with sets of only 6 practice items. Items are ordered by difficulty, and correlations between item difficulties

established separately for people from different backgrounds are nearly all over 0.95, and primarily between 0.99 and 1.00

Section one (verbal section) was placed at the beginning of the instrument because researchers have suggested that it is important for participants to perceive the task as difficult for the stereotype threat effect to manifest (Steele & Aronson, 1995). Section one (verbal) comprised difficult GRE items while items on section two (nonverbal) had a difficulty gradient with some easier items at the beginning.

It is important that some items load on the g-factor because researchers have suggested that the g-factor is highly related to between-group racial differences in academic/IQ performance (Jensen 1998). It is important to have items that measure between-group differences that load on the general factor so that the results may be generalized to the racial IQ gap.

The covariate comprised ACT scores. The ACT is an achievement test typically administered to high school students who are interested in attending college. It highly correlated with the SAT test (a competitor of the ACT), which is also used in ST research as a covariate. The reliability estimate for the ACT published in the 1997 ACT technical Manual is 0.96 (Noble, Roberts, & Sawyer, 2006). Participants' ACT scores were obtained from the registrar's office at each institution. No ACT scores were available for 5 Black and 5 White participants. However, SAT scores were available for these participants. These 10 SAT scores were converted to ACT scores based on tables published by Dorans (1999).

A post-test questionnaire, in addition to collecting various demographic data, solicited self-report of race and presented a deception check item. The deception check item was:

- “What was the purpose of the test you just took? (Circle one):
- (A) It was to determine my personal strengths and weaknesses in verbal ability and intelligence.
  - (B) It was to examine different methods for solving problems
  - (C) It was a standardized test that can be used when applying for scholarships and jobs.
  - (D) It will be used to evaluate my instructors.”

Each choice corresponded to a TD condition as follows: Choice A, STC; Choice B, NSTC; Choice C, CX; and Choice D, distracter.

### Planned Analyses

Data were to be analyzed using 2 x 3 ANOVA design. The analyses were to be computed using the general linear model (GLM), and it was planned to use the SPSS software. Appropriate tests were planned to ensure that all statistical assumptions are met. These assumptions included independence (which was controlled via randomization), normality (which was controlled via sufficient N), and homogeneity of variance (which may be checked by the F-Max Test). Post Hoc analyses were planned. If the interaction of Race and Test Description was detected as significant, then it was planned to examine simple main effects.

### Research Questions and Hypotheses

The primary question to be answered by this study is whether stereotype threat elicitors differentially affect the performance of Blacks and Whites on

standardized cognitive tests such that absence of stereotype threat elicitors will decrease the pre-existing discrepancy in performance between those groups (the racial IQ gap). The following list summarizes the questions that will serve to answer that primary question (see Figure 1 for a graphical representation of these hypotheses):

### Research Question 1

*Is there a significant interaction of the effect of Race and the effect of Test Description on performance on the Cognitive Task?*

A significant interaction indicates the manifestation of the stereotype threat phenomenon. The null hypothesis states that there will not be a significant interaction. The significance of this interaction will be analyzed using the ANOVA F-ratio.

### Research Question 2

*Is the performance of Blacks in the control condition less than the performance of Whites in the control condition?*

A significant difference in the described direction indicates a manifestation of typically occurring racial differences (the racial IQ gap). The null hypothesis states that there will not be a significant difference. The significance of this main effect will be analyzed using the ANOVA F-ratio and visual inspection of the graphed data.

### Research Question 3

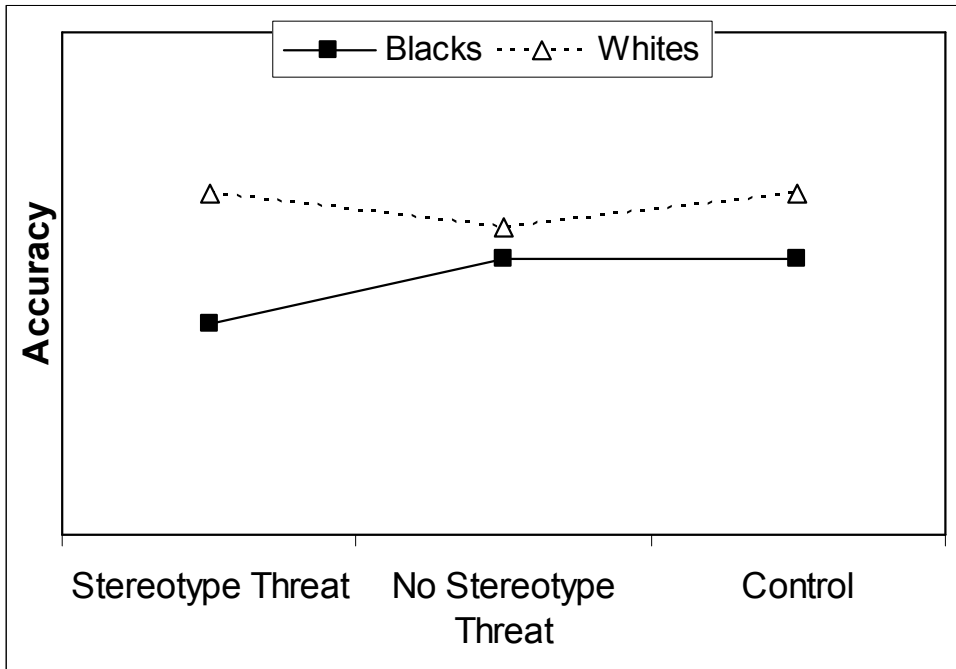
*Is the performance of the Blacks in the NSTC equal to their performance in the control condition but greater than Blacks in the STC?*

Significant relationships in the described directions indicate that stereotype threat elicitors increase the typically occurring racial gap but, when the elicitors are removed, the gap remains the same. The null hypotheses state that no significant relationships will exist. The significance of these main effects will be analyzed using the ANOVA F-ratio and the visual inspection of the graphed data.

### Additional Hypotheses

There remained two additional hypotheses that were based on previous research but were not related to critical research questions. First, across each level of Test Description, it was hypothesized that Whites would perform higher than Blacks. Second, it was speculated that Whites might exhibit slightly depressed performance in the NSTC. This irregularly occurring phenomenon has not been examined before in the literature.

Figure 1  
Research Hypotheses



## CHAPTER IV

### RESULTS

This purpose of this study was to investigate the manifestation of stereotype threat without the use of a covariate and to determine if the manipulation of stereotype threat elicitors can equalize performance on cognitive tasks across racial groups, as is widely reported in the literature. Three primary research questions were articulated to address these issues. To answer these questions, data were analyzed using a two-factor, between subjects analysis of variance design (ANOVA). The independent variables were race (Black, White) and task description (stereotype threat, no stereotype threat, control). The dependent variable was cognitive performance and was measured by accuracy on the cognitive task. Scores on the verbal section of the cognitive task were omitted from analysis due to a floor effect, which is discussed below. Thus, only nonverbal performance was analyzed to answer the research questions. A chi square analysis was performed on the manipulation check data to examine whether the task description deception was successful. Also, to permit a comparison of this analysis (ANOVA) to traditional stereotype threat research (ANCOVA), data were re-analyzed with ACT scores as a covariate. The

Statistical Package for the Social Sciences version 12.0 was used to conduct statistical analyses of the data.

### Sample

Data from 160 participants were collected for the analyses. Although the design prescribed equal cell sizes, obtained cell sizes were not equal (see Table 3). This was assumed to be the result of chance. Therefore, the unique sums of squares approach (default SPSS GLM approach) was used to adjust for unequal  $n$  in the analyses of variance and the analyses of covariance.

### Assumptions

The planned statistical analyses each required specific assumptions to be met. Table 2 presents each of these assumptions and indicates whether they were met. The analysis of variance design applied to the nonverbal performance data was the primary analysis for this study and was used to answer the research questions. For that analysis, all statistical assumptions were met. The chi square analysis was applied to the deception check data, and all statistical assumptions of the design were met. Three outliers ( $z > 3$ ) were present in the verbal performance data. These scores were not a result of measurement error and are assumed to be legitimate scores. Because the verbal data were omitted from the primary analysis due to a floor effect, no data transformations or omissions were applied. Heterogeneity of regression slopes was detected for the analysis of covariance of nonverbal performance data, which was a predicted



violation. However, the analysis was performed despite this violation to permit scrutiny of the design, which is used in most stereotype threat research without explicit regard for statistical assumptions.

Table 2  
*Conformity of Data to Statistical Assumptions by Analysis*

Assumption	Nonverbal		Verbal		Deception
	ANOVA	ANCOVA	ANOVA	ANCOVA	Chi Square
Independence	Yes	Yes	Yes	Yes	Yes
Normality	Yes	Yes	Yes	Yes	
Homogeneity of Variance	Yes	Yes	Yes	Yes	
Absence of Outliers	Yes	Yes	No	No	
Linearity of Regression		Yes		Yes	
Homogeneity of Regression		No		Yes	
Reliability of Covariate		Yes		Yes	

#### Analyses on Nonverbal Performance

Nonverbal performance data were analyzed using a two-factor, between subjects, 2x3 analysis of variance design to answer the three primary research questions. Independent variables consisted of race (Black and White) and task description (stereotype threat, no stereotype threat, control). All statistical assumptions were met. Descriptive statistics are presented in Table 3, and results of the analysis are presented in Table 4 and Figure 2. The ordinal interaction effect of race and task description reached marginal statistical significance,  $F(2, 154) = 2.97, p = 0.054$  (see Figure 2). The partial Eta squared was 0.037, indicating that the interaction accounted for 3.7% of the total (effect + error) variance. A main effect of race,  $F(1, 154) = 22.01, p = 0.000$ , and a main effect of task description,  $F(2, 154) = 5.17, p = 0.007$ , were also detected.

For post-hoc analysis, cell means were inspected by graph to answer research questions (Figure 2). For descriptive purposes, significance levels and strengths of association ( $\eta_p^2$ ) for 11 independent post-hoc comparisons were calculated using SPSS-GLM univariate (see Table 5). Interpretation of such values remains controversial (Keppel, 1991). The significance levels were not adjusted for the inflation of the probability of Type I error resulting from multiple comparisons. Disregard for this increase in familywise error is common practice in psychology, according to Keppel (1991), especially when the omnibus test is significant. A number of techniques are available for those interested in controlling for error inflation. Using a Bonferroni adjustment, significance values reported in Table 5 may be evaluated at the  $\alpha = 0.0045$  level to insure that familywise error does not exceed  $\alpha = 0.05$ . Notation is used in Table 5 to differentiate values that reach unadjusted significance from those that remain significant after the Bonferroni adjustment. Keppel (1991) recommends that values which are only significant prior to the Bonferroni adjustment be considered “suggestive of significance.” Only recognizing values that are significant at the Bonferroni-adjusted level can lead to significant loss of power. Furthermore, overconcern with Type I error can lead to unacceptable inflation of Type II error.

Several characteristics of this study are unique and support its categorization as exploratory, contraindicating a Bonferroni adjustment or other conservative procedures. First, the participants in this study were recruited from less selective universities and are generally less academically capable than those in previous research. Second, such a difference in ability might be

confounded by correlation with variables known to influence stereotype threat, such as domain identification. Third, this is the first study to use a control condition to compare the conventional STC and NSTC to a control that simulates a typical standardized testing milieu. Fourth, this study is the first to not use a covariate or other statistical manipulation of data. Fifth, the purpose of this study is to inform future research, not practice or application. Thus, because novel characteristics in this study render hypotheses speculative and because the purpose is to guide continued research, care should be given to not neglect possible significance in results.

Reanalyzing the nonverbal data using analysis of covariance permits a comparison of this study's design to traditional stereotype threat research design and illustrates the misleading nature of using a covariate in such studies. Thus, nonverbal performance data were also analyzed using a two-factor, between subjects, 2x3 analysis of covariance. Independent variables remain the same as the previous analysis of variance. Participants' ACT scores were entered as the covariate. As predicted, heterogeneity of regression slopes was detected. All other statistical assumptions were met. Descriptive statistics are presented in Table 6, and results of the analysis are presented in Table 7 and Figure 3. After covarying for ACT scores, the interaction effect of race and task description was disordinal and reached statistical significance,  $F(2, 153) = 3.54, p = 0.031$  (see Figure 3). The partial Eta squared was 0.044, indicating that the interaction accounted for 4.4% of the total (effect + error) variance. The main effect of task description also reached statistical significance,  $F(2, 153) = 5.25, p = 0.006$ .

After covarying for ACT scores, the main effect of race was no longer significant,  $F(1, 153) = 0.88, p = 0.349$ . Thus, using the ACT covariate statistically eliminated the significant discrepancy in performance between racial groups on the nonverbal section.

Table 3  
*Descriptive Statistics for ANOVA on Nonverbal Performance*

Race	Task Description	Std.		Minimum	Maximum	n
		Mean	Deviation			
Black	Stereotype threat	59.8%	19.3%	34.6%	94.7%	27
	No stereotype threat	42.2%	21.5%	30.0%	90.4%	29
	Control	58.1%	24.8%	45.5%	95.8%	24
	Total	52.9%	23.1%	30.0%	95.8%	80
White	Stereotype threat	64.8%	19.6%	29.2%	94.7%	28
	No stereotype threat	65.7%	18.6%	13.3%	88.2%	25
	Control	74.3%	15.2%	16.7%	94.7%	27
	Total	68.3%	18.2%	13.3%	94.7%	80
Total	Stereotype threat	62.4%	19.4%	29.2%	94.7%	55
	No stereotype threat	53.1%	23.3%	13.3%	90.4%	54
	Control	66.7%	21.7%	16.7%	95.8%	51
	Total	60.6%	22.1%	13.3%	95.8%	160

Table 4  
*Analysis of Variance for Nonverbal Performance*

Source	df	F	p	$\eta_p^2$	observed power
Race (R)	1	22.01*	0.000	0.125	0.997
Task Description (T)	2	5.17*	0.007	0.063	0.821
R x T	2	2.97	0.054	0.037	0.570
Error	154				

Note. \* $p < .05$

Table 5  
*Post-Hoc Analyses on Nonverbal Performance Data*

Group Analyzed	Compared Means	significance	$\eta_p^2$
Black (B)	Between TD groups	0.006 **	0.125
Black	ST v NST	0.002 ***	0.161
Black	ST v CX	0.078 *	0.002
Black	NST v CX	0.016 **	0.109
White (W)	Between TD groups	0.105 *	0.057
White	ST v NST	0.863	0.001
White	ST v CX	0.050 **	0.070
White	NST v CX	0.073 *	0.063
Stereotype threat (ST)	B v W	0.351	0.016
No stereotype threat (NST)	B v W	0.000 ***	0.258
Control (CX)	B v W	0.007 **	0.141

*Note.* Group abbreviations are denoted in the Group Analyzed column. The abbreviation TD indicates task description.  $\alpha_{\text{planned}} = 0.05$ ;  $\alpha_{\text{familywise}} = 0.0045$ . \* marginally significant without familywise (FW) adjustment; \*\* significant without FW adjustment; \*\*\* significant with and without FW adjustment.

Table 6  
*Descriptive Statistics for ANCOVA on Nonverbal Performance*

Race	Task Description	Actual Mean	Adjusted Mean	Adjustment to Mean	ACT Mean	n
Black	Stereotype threat	59.8%	64.1%	+4.3%	16.96	28
	No stereotype threat	42.2%	48.5%	+6.3%	15.76	25
	Control	58.1%	63.7%	+5.6%	16.21	27
	Total	52.9%	58.8%	+5.9%	16.30	80
White	Stereotype threat	64.8%	57.3%	-7.5%	23.79	27
	No stereotype threat	65.7%	61.0%	-4.7%	22.16	29
	Control	74.3%	70.4%	-3.9%	21.70	24
	Total	68.3%	62.9%	-5.4%	22.58	80
Total	Stereotype threat	62.4%	60.7%	-1.7%	20.44	55
	No stereotype threat	53.1%	54.8%	+1.7%	18.72	54
	Control	66.7%	67.0%	+0.3%	19.12	51
	Total	60.6%	60.8%	+0.2%	19.44	160

*Note.* Adjustment was calculated using ACT scores as the covariate

Table 7  
*Analysis of Covariance for Nonverbal Performance*

Source	df	F	p	$\eta_p^2$	observed power
ACT	1	11.37*	0.001	0.069	0.918
Race (R)	1	0.88	0.349	0.006	0.154
Task Description (T)	2	5.25*	0.006	0.064	0.827
R x T	2	3.54*	0.031	0.044	0.652
Error	153				

*Note.* ACT is the covariate. \* $p < .05$

Figure 2  
*Mean Performance by Racial Groups on Nonverbal Cognitive Task*

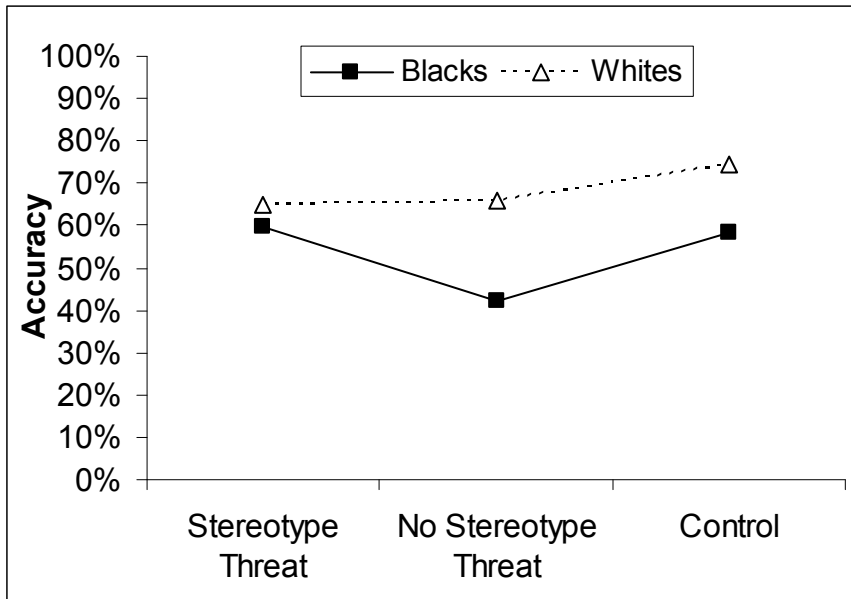
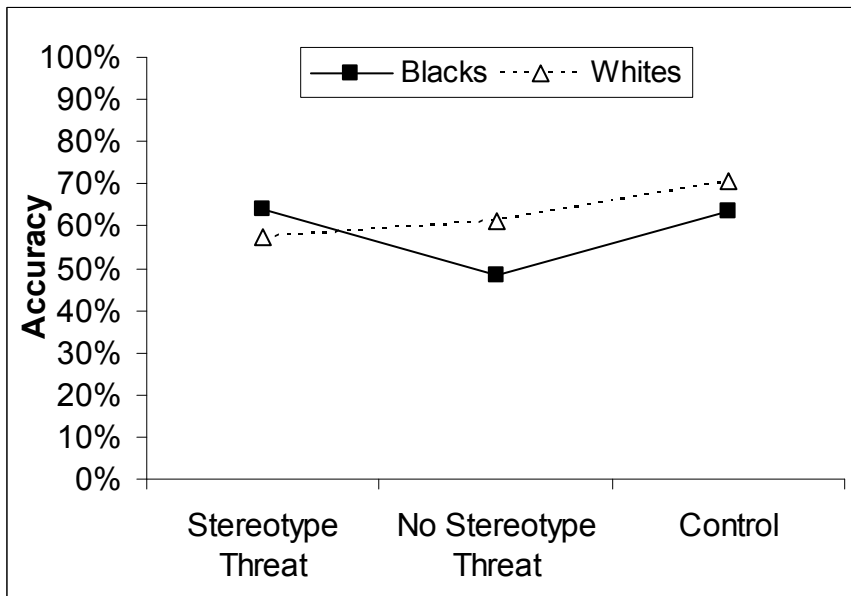


Figure 3  
*Adjusted Mean Performance by Racial Groups on Nonverbal Cognitive Task*



Note. Means in Figure 3 are adjusted by ACT covariate.

## Analyses on Verbal Performance

Performance on the verbal section of the cognitive task was omitted from the primary analyses and not applied to the research questions because a floor effect manifested across all cells. A floor effect is a condition by which test items were too difficult to differentiate scores in the bottom portion of the distribution. The size of the portion of the distribution that remains undifferentiated corresponds to excessiveness of the test difficulty. Often, undifferentiated scores accumulate in the bottom tail of the distribution resulting in a positively skewed distribution. A test that is so difficult that no participant can discern the correct response for any item would completely fail to measure any difference between participants' performance. This situation will be referred to in this report as a "complete floor effect." With a complete floor effect, the entire distribution is undifferentiated.

The grand mean for performance on the Verbal Section was 23.2%. Because items presented only 5 response choices, random responding alone would produce a mean performance of 20%. Thus, a complete floor effect would be indicated by a mean performance of 20%. Figure 4 presents the unadjusted means for each condition. It is clear that participants, especially Black participants (mean performance = 19.9%), performed no better than would random responding. This inability of the instrument to detect differences in participants' performance results in nullification of the effects of the experimental manipulations and prohibits an examination of the effect of stereotype threat elicitors on participants' performance. In prior research and in the nonverbal data

from this study, the task description has had a substantial influence on the performance of Blacks. However, because of the floor effect in the verbal section, the mean verbal performance of Blacks in this study were all near the 20% random response level (see Table 8) and there was no main effect of task description on performance for Blacks,  $F(2,77) = 0.514, p = .60$ .

Often floor effects result in non-normal distributions. These distributions often appear positively skewed, because participants who would score in the bottom portion of the distribution are not differentiated. However, the use of multiple choice items can mask this characteristic of the floor effect because chance can create a normal distribution around the random response level. To clarify this point, consider the following two examples. Test A requires participants to recall responses to items with no provided response choices (i.e., fill in the blank), but manifests a complete floor effect whereby all participants do not know any correct responses. Test A data would have a highly kurtotic non-normal distribution with a mean of 0%. Test B provides two response choices per item, permits guessing, but also manifests a complete floor effect whereby all participants do not know any correct responses. For test B, participants responses are no more likely to be correct than responding by flipping a coin. Nevertheless, many participants will guess correctly on some items. Indeed, with an adequate sample size, one would expect a normal distribution, with a mean of 50%. Similarly, because the verbal section employed multiple choice items like hypothetical “test B”, normality of the distribution cannot be used to rule out the



floor effect. The distribution of verbal performance scores for all participants was characterized by a skewness of .670 and kurtosis of 1.671.

Although the floor effect reasonably precludes the application of verbal data to the research questions, results of analyses are presented to fully describe the verbal dataset for readers who are skeptical of the floor effect. Additionally these analyses provide an additional illustration of the misleading nature of covariate used in this type of study.

The data were analyzed using a two-factor, between subjects, 2x3 ANOVA design. Then, data were analyzed using a two-factor, between subjects, 2x3 ANCOVA design (with ACT scores as a covariate), as is traditionally used in stereotype threat research. With alpha set at 0.05, the interaction effect was insignificant for both the ANOVA,  $F(2,154) = 1.590$ ,  $p = .207$ , and the ANCOVA,  $F(2,153) = 1.145$ ,  $p = .321$ . This confirms that the floor effect masked detection of the stereotype threat effect on verbal performance.

The ANOVA yielded a significant main effect of race,  $F(1,154) = 18.742$ ,  $p=0.000$ . However, when the ACT covariate was employed, the means for both racial groups were equal (23%). Thus, as with the nonverbal data, using the ACT covariate statistically eliminated the significant discrepancy in performance between racial groups on the verbal section. Table 8, Table 9, Figure 4, and Figure 5 present descriptive statistics and results for the analyses on verbal performance.

Table 8  
*Descriptive Statistics for Analyses of Verbal Performance*

Race	Task Description	Actual Mean	Adjusted Mean	Adjustment to Mean	ACT Mean	N
Black	Stereotype threat	20.4%	23.1%	+2.7%	16.96	28
	No stereotype threat	18.5%	22.4%	+3.9%	15.76	25
	Control	20.9%	24.3%	+3.4%	16.21	27
	Total	19.9%	23.3%	+3.4%	16.30	80
White	Stereotype threat	30.6%	26.0%	-4.6%	23.79	27
	No stereotype threat	24.5%	21.5%	-3.0%	22.16	29
	Control	24.5%	22.1%	-2.4%	21.70	24
	Total	26.6%	23.2%	-3.4%	22.58	80
Total	Stereotype threat	25.6%	24.5%	-1.1%	20.44	55
	No stereotype threat	21.3%	22.0%	+0.7%	18.72	54
	Control	22.8%	23.2%	+0.4%	19.12	51
	Total	23.2%	23.2%	0.0%	19.44	160

*Note.* Adjustment was calculated using ACT scores as the covariate

Table 9  
*Analyses of Verbal Performance*

	Source	df	F	p	$\eta_p^2$	observed power
ANOVA	Race (R)	1	18.742*	0.000	0.108	0.990
	Task Description (T)	2	2.525	0.083	0.032	0.500
	R x T	2	1.590	0.207	0.020	0.333
	Error	154				
ANCOVA	ACT	1	20.165*	0.000	0.116	0.994
	Race (R)	1	0.002	0.965	0.000	0.050
	Task Description (T)	2	1.012	0.366	0.013	0.224
	R x T	2	1.145	0.321	0.015	0.249
	Error	153				

*Note.* ACT is the covariate. \*p<.05

Figure 4  
Mean Performance by Racial Groups on Verbal Cognitive Task

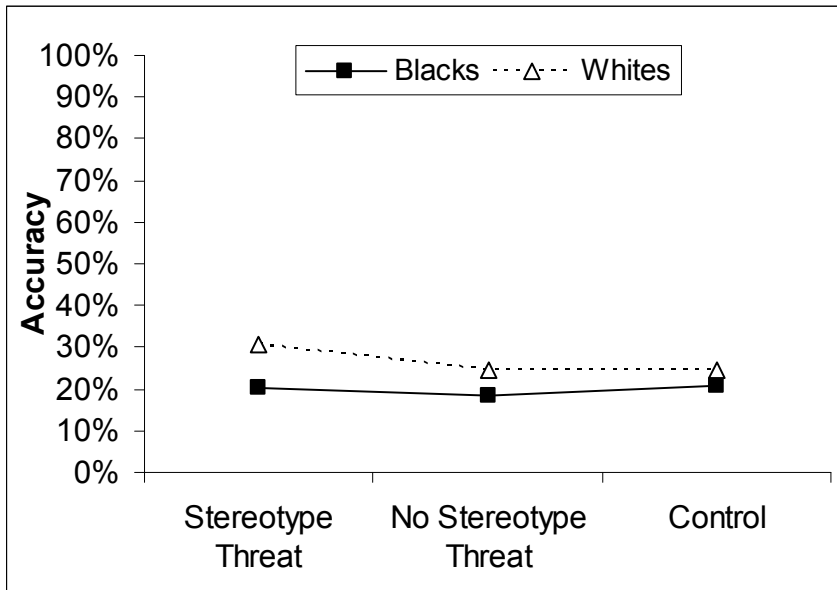
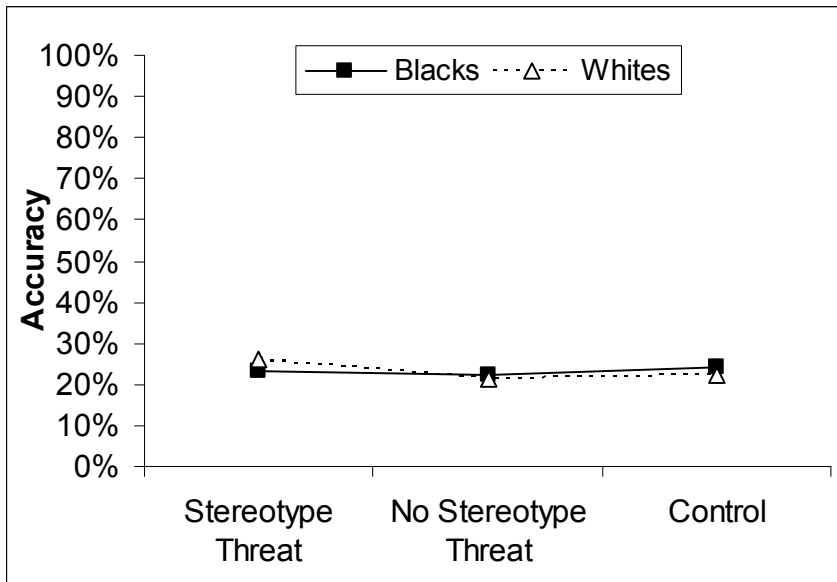


Figure 5  
Adjusted Mean Performance by Racial Groups on Verbal Cognitive Task



## Analysis of Deception Check

One multiple choice item in a post-test questionnaire measured whether the task description deception was successful by prompting each participant to identify the purpose of the study from four response choices. One choice corresponded to each task description condition (stereotype threat, no stereotype threat, control) and one item served as a distracter. If the deception was successful, the modal response choice for each level of task description should vary significantly. Data were analyzed using the chi squared statistic, and all assumptions of the design were met. The analysis indicated that the differences were significant,  $\chi^2 (4, N = 160) = 237.981, p = 0.000$ . Also, there was no difference between races,  $\chi^2 (4, N = 160) = 1.172, p = 0.556$ , indicating that the deception was equally effective for both Black and White participants. Table 10 presents the percentage of participants' responses for each response choice across each experimental condition. It should be noted that endorsement of the incorrect item by an individual participant does not necessarily indicate that the participant did not believe the deception. This is due to possible differences in participants' interpretation of the questionnaire deception item. However, the significant difference in the distribution of responses across conditions is a strong indicator that the task description manipulations were salient to the participants and affected their overall perception of the purpose of the task.

Table 10  
*Percentage of Participants' Responses by Experimental Condition*

Race	Study Condition	Response on Deception Check Item			
		Stereotype threat	No stereotype threat	Control	Distracter
Black	Stereotype threat	96%	4%	0%	0%
	No stereotype threat	14%	86%	0%	0%
	Control	13%	4%	83%	0%
White	Stereotype threat	100%	0%	0%	0%
	No stereotype threat	16%	84%	0%	0%
	Control	11%	0%	89%	0%

## Research Questions

### Research Question 1

*Is there a significant interaction of the effect of Race and the effect of Test*

*Description on performance on the Cognitive Task?*

The interaction effect of race and task description reached marginal statistical significance,  $F(2, 154) = 2.97, p = 0.054$ . A significant interaction indicates the manifestation of the stereotype threat phenomenon. The omission of the ACT covariate from the conventional stereotype threat design increased the error term, resulted in a loss of power, and likely accounts for the elevated alpha level of this interaction relative to previous ST studies. Because the F-ratio reached marginal significance in the planned analysis of variance and reaches substantial significance in the traditional analysis of covariance, it can be reasonably inferred that a phenomenon, analogous to that which has been observed in previous stereotype threat research, did manifest in this sample. The null hypothesis for the first research question was rejected.

### Research Question 2

*Is the performance of Blacks in the control condition less than the performance of Whites in the control condition?*

A significant difference in the described direction indicates a manifestation of typically occurring racial differences. Mean Black nonverbal performance in the control condition was 58.1%. Mean White nonverbal performance in the control condition was 74.3%. This yields a discrepancy of 16.2 percentage points. Relative to the total variability in the control condition of this non-proportional sample, this is a significant 0.75 standard deviation difference. Using the distribution of scores of White participants, which is more generalizable to the White-majority distributions used in racial differences research, yields a 1.1 standard deviation discrepancy. This closely approximates the 1 standard deviation discrepancy commonly reported in racial differences research. The null hypothesis for the second research question was rejected.

### Research Question 3

*Is the performance of the Blacks in the no stereotype threat condition equal to their performance in the control condition but greater than Blacks in the stereotype threat condition?*

Significant relationships in the described directions indicate that stereotype threat elicitors exacerbate the typically occurring racial gap, but manipulation (i.e. removal) of the ST elicitors does not eliminate the racial gap. Although the level of task description affected the observed performance of Blacks, the influence of the ST elicitors (i.e., task description levels) on performance of Black participants in the no stereotype threat condition and the stereotype threat condition was opposite of the prediction. Whereas, the presentation of stereotype threat elicitors depressed Blacks' performance in prior research, the absence of stereotype threat elicitors depressed Blacks' performance in this sample. Despite this unpredicted stereotype threat effect reversal, Blacks' performance in neither the stereotype threat condition nor the no stereotype threat condition was significantly greater than their performance in the control condition. As predicted, stereotype threat elicitors had a significant influence on Blacks' performance, but neither the presentation nor removal of these elicitors served to equalize performance with White participants in the control condition. Additionally, neither the presentation nor removal of the stereotype threat elicitors served to improve Blacks' performance relative to their performance in the control condition (see Figure 2, Table 3, and Table 5). The null hypothesis for the third research question was rejected, but the predicted relative levels of two of the three conditions (Stereotype threat and No stereotype threat) were reversed.

### Additional Hypotheses

Two additional hypotheses were proposed based on previous research but were not related to critical research questions. First, across each level of Test Description, it was hypothesized that mean White performance will be higher than mean Black performance. Although this pattern of scores is masked by a covariate in analyses of covariance (see Figure 3 and Figure 5), analysis of variance of unadjusted scores reveals a pattern of scores in congruence with this prediction (see Figure 2 and Figure 4). Second, based on previous research, it was predicted that Whites' performance will not vary across levels of Test Description. Mean White performance did not appear to vary significantly between the stereotype threat and the no stereotype threat conditions. These were similar to conditions employed in previous research. However, a control condition was employed which was unique to this study. Mean White performance in the control condition was higher than mean White performance in either other condition (see Figure 2, Table 3, and Table 5).

### Limitations and Future Research

This study exhibits a number of limitations which must be considered. These include issues of population representation, confounding variables, and instrumentation flaws. While these limitations indicate that caution is necessary for some specific interpretations of this data, the general conclusion (i.e., ST elicitors cannot account for the racial gap) remains strongly supported.



Although an attempt was made to obtain a more representative sample than previous ST research, the obtained sample was composed only of college students. This narrow demographic set limits inferential freedom. Nevertheless, the conclusion drawn from the data (discussed in Chapter 5) is that more caution is necessary when inferring ST research results. Although the obtained sample was a biased selection, it was discrepant enough from previous ST research to show that ST elicitors have different effects on different samples; that the ST effect is not universal. Thus, the selection bias which is evident in all ST research does impact the ST elicitors' effect, and caution is justified.

Participants in this study were assigned to racial groups based on self-report. The validity of such measurement is questionable, but this procedure is consistent with other ST research. The racial group variable in this study is confounded by site (i.e., university). This compromise was necessary to acquire an adequate Black sample that was relatively representative. It was previously discussed that education has only a small impact on performance on IQ tests (nonverbal task), so there is little reason to suspect that site (one to two years of education for participants) independently influenced performance on the nonverbal test. There is no reason to suspect that the site variable interacted with the level of Task Description to influence performance, although this possibility cannot be empirically ruled out. While the site variable is an easily identified confound, the race variable itself may be a proxy for a constellation of confounding variables (e.g., SES, cultural values, etc.), and it is unclear whether the race variable was any more confounded in this study than previous ST

studies. Because participants were assigned by site to racial groups, sampling procedure was not truly random. The racial minority participants at the respective institutions (Blacks at OSU; Whites at LU) were excluded because some previous ST research has demonstrated that minority membership can be a ST elicitor and their inclusion would have confounded the TD variable.

The analyses for this study were limited because of a floor effect of the verbal task. This was a consequence of using the GRE items previously used in ST research with the less selective sample of this study. Unfortunately, the order of task presentation (verbal and nonverbal) was not counterbalanced. Thus, exposure to the difficult verbal items may have influenced subsequent performance on the nonverbal task across all experimental conditions, although there is no substantial theoretical or empirical support to indicate such an effect. During testing, participants received no feedback, so it is unclear whether or not they were aware of their verbal performance. Additionally, it was hypothesized that exposure to a difficult task is necessary for the ST effect to manifest, and the difficulty gradient of the nonverbal task may have reduced the measured interaction effect if the tasks were counterbalanced.

The stereotype threat condition for this study presented a number of ST elicitors (e.g., task description, racial priming) and no attempt was made to differentiate the effect of individual elicitors. Also, because nonverbal performance of Blacks was approximately equal between the control condition and the ST condition, one might speculate that both conditions manifested the ST effect and optimal Black performance was not measured across any of the

three experimental conditions. However, the ST effect is operationally defined in research by the effect on cognitive performance of specified elicitors, and no ST elicitors were presented in the CX or the NSTC. Although it cannot be determined whether some unmeasured contextual variable (i.e., ST elicitor) depressed Black (or White) performance across all experimental conditions, it is certain there was no effect of the specified ST elicitors (i.e., ST effect) in the CX or the NSTC as defined by ST research because such variables were controlled in these conditions. More importantly, it is reasonable to infer from these results that racial cognitive performance cannot be equalized in the general population by manipulating these ST elicitors. Inferring that Black performance in the CX was less than optimal and depressed by unknown contextual variables is entirely speculative and devoid of any empirical support from this or other research.

Some readers may criticize this report's acknowledgement of Hereditarian authors, and perceive it as a limitation to this study's interpretation. However, the design, execution, analysis, and interpretation of this study are not dependent in any way on historical research, hereditarian or otherwise. This study was empirically based, and interpretations are objectively rendered. While the literature review of this report did describe the work of hereditarians (and other perspectives), it was done to provide contextual understanding. It should be noted that some of these hereditarians were renowned psychologists. For example, Richard Herrnstein, a Harvard University professor of psychology who studied with B.F. Skinner, was a highly productive behaviorist before publishing

*The Bell Curve*. Some research studies indicate that his work represented the opinion of many experts at that time, and it cannot be omitted from review because it violates personal values. Additionally (as discussed in Chapter II), the work of such authors (1) is unpopular among many editors, which makes peer reviewed articles difficult to publish; and (2) is often the only group that attempts to examine specific characteristics of the racial IQ gap.

Rather than providing strong support for any particular conclusion, this study significantly weakens the interpretation of previous research. The study and critical review presented in this report designates specific weaknesses in ST theory and identifies necessary directions for future research. Such research should (1) be in the school setting using school-aged children, (2) not covary out variance related to the group differences of interest, (3) use pretest/posttest measures or control groups to show the direction of the ST effect unmasked by statistical manipulation, (4) use g-loaded tests, (5) involve students across a reasonable range of functioning, and (6) measure demographic variables known to be related to the group differences of interest.

## CHAPTER V

### DISCUSSION

Claude Steele and Joshua Aronson (1995) identified a unique psychological phenomenon when they showed that subtle manipulation of specific contextual variables (stereotype threat elicitors) can differentially affect performance of racial groups on standardized tests. While the idea that contextual variables can differentially affect the performance of groups was not novel, the supposition of many ST researchers that these manipulations can eliminate the racial gap was profound. There are several important implications for stereotype threat theory which are evident from previous research reviewed in this report and are empirically supported by the results of this study: (1) the typical ST phenomenon occurs in a restricted segment of the Black population (high achievers); (2) the adverse effect of the ST elicitors on Black performance is significantly attenuated when real world contexts are simulated (see also: Brown & Day, 2006); (3) the ST elicitors only exacerbate the typical racial gap in performance on standardized test and cannot not be used to diminish it; (4) Using a covariate that is highly correlated with performance on a standardized test (i.e., another standardized test) can reduce the racial gap via statistical manipulation.

Objective review of the available literature revealed a number of weaknesses of ST theory. Critics could consider stereotype threat theory to be a fallacy of logic, *circulus in demonstrando*: Blacks tend to perform lower in IQ tests because of negative stereotypes about Blacks performing lower on IQ tests. Regardless of such violations of logic, this literature review and the study in this report delineate at least five arguments that are each independently sufficient to reject ST theory as a likely explanation for the depressed performance of Blacks on standardized tests: (1) the statistical mischaracterization, (2) the Spearman hypothesis, (3) unrepresentative research samples, (4) unthreatened populations, (5) and psychometric properties of the standardized tests. These arguments are supported by a convergence of findings from this study, previous ST research, and decades of research on racial differences.

#### Mischaracterization

Rudimentary statistical knowledge (i.e., knowing the function of a covariate) should have been sufficient to empower researchers to recognize the error of previous stereotype threat research interpretations. Statistics are employed to provide an objective evaluation of data, and there is little ambiguity in the results of most ST research. A preponderance of ST research has shown that ST elicitors only exacerbate the racial gap, but many ST researchers' interpretations are opposite of this requisite conclusion (Sackett et al., 2004a). ST theorists usually hypothesize that the manipulation of ST elicitors can be used to elevate the mean performance Blacks on standardized tests to be equivalent

to the typical mean performance of Whites. The empirical results of this study strongly support the rejection of this mischaracterization. Certainly in the sample obtained for this study, ST elicitors were manipulated successfully (as shown by analysis of deception check data and the main effect of TD), and although such manipulations did affect performance of Black participants, racial group performance was not equalized under any condition. This was demonstrated with relatively representative samples and without covarying or employing any other statistical manipulation of the data.

The analyses of the nonverbal and verbal data from this study demonstrate that using the ACT covariate, as most previous ST researchers have done, statistically equalizes between-group performance, a priori. The analyses of variance (no covariate) yielded significant main effects of race for both the nonverbal data,  $F(1, 154) = 22.01, p = 0.000$ , and verbal data,  $F(1, 154) = 18.742, p = 0.000$ . Additionally, the significant interaction effects of these analyses of variance were ordinal, whereby Whites scored higher than Blacks across all conditions. When a covariate was entered in the analyses, there was no longer a significant main effect of race for the nonverbal data,  $F(1, 153) = 0.88, p = 0.349$ , or the verbal data,  $F(1, 153) = 0.002, p = 0.965$ . Thus, the overall racial discrepancy in performance disappeared. When Task Description means were adjusted by the covariate, Blacks appeared to outperform Whites in the STC of the nonverbal data and in the NST and CX of the verbal data, although racial differences in the verbal data were quite small. Equalizing the performance of racial groups in most stereotype threat studies is not an effect of

the manipulation of stereotype threat elicitors (e.g., task descriptions, racial priming), but is a result of a statistical manipulation (covariance). The precise influence of the covariate on racial group means is presented in Table 6. For example, the covariate raised the mean nonverbal performance of Blacks by 5.9% while simultaneously lowering the mean nonverbal performance of Whites by 5.4%. This statistical manipulation diminished the observed racial gap (15.4%) in nonverbal performance by 73.4%. As discussed in Chapter II, one study initially appeared to demonstrate equalization of group performance without a covariate (Brown & Day, 2006). However, close scrutiny revealed that these authors equalized groups a priori, not with a covariate, but with an unrepresentative sample.

The use of a covariate like the ACT was also criticized for the possible violation of statistical assumptions. Heterogeneity of the regression coefficients was detected in the nonverbal data of this study. The distribution of the verbal data was altered by a significant floor effect, so investigating the violation of distribution assumptions in the verbal data is not useful.

Using empirical evidence, this study supports the criticisms, first delineated by Sackett and colleagues (2004a), that previous interpretations of stereotype threat ANCOVA designs are inappropriate. Knowing how to increase or exacerbate the racial gap by manipulating ST elicitors has questionable utility. Although covarying on prior performance can prevent the measurement of between-group racial discrepancies in test data, such a process has no real utility and is merely an artifact of misused statistical analyses. Indeed, most



standardized tests already overpredict Blacks' performance on future criteria (Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997; Noble, 2003), and statistically elevating Blacks' scores will further diminish the instruments' validity.

An additional misinterpretation was suggested by the results of this study. Mean White performance in the control condition, which was not used in previous ST research, was higher than mean White performance in either other condition (STC or NSTC)(see Figure 2, Table 3, and Table 5). This finding suggests that mean White performance in previous ST studies (which only use conditions analogous to the STC and NSTC) may not be generalizable to mean White performance in real-world settings, which is estimated by the CX in this study. Specifically, previous ST research may use conditions that substantially depress White performance. Even if previous ST researchers had successfully elevated mean Black performance to the level mean White performance in their experimental conditions (which they did not), the scores of both groups would have likely been lower than the mean performance of Whites in a real-world setting. This issue of generalizability is independent of the problems of sample representation discussed later in this chapter.

### ST Predictions

Decades of research from dozens of unaffiliated researchers across different countries have illuminated several specific characteristics of the racial gap. These characteristics cannot be ignored and must be predicted by any theory that purports to account for the racial gap. ST theory not only fails to

account for these findings, but implicit predictions of ST theory explicitly contradict the historical data. Discrediting such an enormous volume of research is a prodigious burden to bear for ST theorists. Thus far, ST theorists have avoided acknowledging these known characteristics and have avoided a discussion of the incompatibility between ST theory and other racial differences research. Yet, onus probandi lies clearly with stereotype threat theorists, especially those who suggest its applicability in schools

### The G Factor

The racial gap in performance on standardized cognitive tests is most apparent on highly g-loaded tasks (see Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). This consistent finding is frequently referred to as the Spearman Hypothesis. Most stereotype threat researchers use tasks, such as verbal achievement tests, that are not highly g-loaded (see Brown, 2001; Davis III, 2000; Steele & Aronson, 1995). The few exceptions include McKay (McKay 1999; McKay et al., 2002; McKay, Doverspike, Bowen-Hilton, & McKay, 2003) who found mixed results while covarying and Brown and Day (2006) who used an inappropriate sample. Because ST elicitors affect Black performance regardless of g-loading and because there is no proposed or apparent explanation for the Spearman Hypothesis using the ST theory, researchers should be skeptical of ST theory's ability to explain the racial gap.

## ST Threatened Populations

Many ST researchers specify prerequisite participant characteristics which are necessary for the manifestation of the ST effect. These include high domain identification, high group identification, knowledge of the stereotype, and stereotype salience. The stereotype threat model (i.e., specifying that ST elicitors account for racial differences) focuses on high achieving, college-aged students (see Aronson, Lustina, & Good, 1999; see also Brown, 2001; Davis III, 2000; McKay, 1999; Steele & Aronson, 1995) and depends on participant knowledge of stereotypes of Black intellectual inferiority. However, it is known that the racial gap persists across the entire range of functioning (not just high-achievers) and manifests in children as early as 3 years old (Levin, 1997). The ST model's implicit prediction that the gap only manifests in older, highly achieving individuals is in conflict with racial differences research. School-aged children, who are still developing an identity and who are likely unaware of racial stereotypes during testing, should not be affected by ST elicitors. These school-aged children must routinely perform on standardized tests and are of great concern to many practitioners who wish to intervene on the depressed performance of Black children. Although the ST model is not capable of accounting for the racial gap in such a demographic, these students are the focus of some of the stereotype threat theorists who publish applications of the theory (see Aronson, 2002; Steele, Spencer, & Aronson, 2002).

Additionally, the stereotype threat model implicitly predicts that the racial gap should be isolated to cultures in which there is a well-known stereotype

against African-American intellectual ability. It is discussed in stereotype threat literature that such a stereotype is well-known in the US and some other European countries (see Croizet, Desert, & Dutrevis, 2001; Steele, Spencer, & Aronson, 2002). However, decades of research show that the gap manifests in numerous countries and cultures, some of which comprise a Black majority (Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). It would be reasonable to assume that Blacks (and especially Black children) who live in countries where Blacks comprise the majority of the population may not feel persistently threatened by Western stereotypes.

Foreign Black majority populations are not the only racial demographic whose depressed performance violates ST model predictions. ST elicitors have been shown to depress Whites' performance when comparison to Asians is implied. Again, the research focus has been on older, high-achievers. Traditional racial differences research has demonstrated that White mean performance on standardized tests is indeed depressed relative to East Asians and Jews. However, these racial gaps are not confined to older, high achievers (Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997). To account for these racial gaps, the ST model (as applied to Blacks) would implicitly predict that Whites who exhibit the relatively depressed performance (1) highly identify with academic achievement, (2) are cognizant of their racial group membership, (3) are aware of a stereotype about their intellectual inferiority, and (4) that these issues are salient during test administration. In other words, White students in American schools are persistently threatened by perceived stereotypes about their

intellectual ability compared to Asian children. The apparent absurdity of these predictions facilitates recognition of the error in ST theorists' reasoning. There are other such unlikely predictions. For example, following the ST model, one must infer from ST research (see Stone, Lynch, & Sjomeling, 1999) that Blacks tend to outperform Whites on some athletic tasks because White athletes are perpetually concerned with stereotypes about their physical inferiority.

### Generalizability

As discussed, the ST model assumes that ST research findings derived from populations of academic high-achievers are generalizable to the general population. This study provides empirical evidence that the ST effect does not extend across the continuum of academic achievement or intellectual functioning, and also suggests that the effect of ST elicitors may be reversed in populations comprised of average individuals (instead of only high-achievers).

In one of Steele and Aronson's 1995 studies, the reported mean SAT score for Blacks was 603. The publisher of the SAT reported that, in 1996, the national mean SAT score for Blacks was 434, with a standard deviation of 99. Thus, Steele's and Aronson's sample of Black students exhibited an SAT mean that was 1.7 standard deviations above the national Black mean and was even 0.74 standard deviations above the national mean of Whites (see College Board, 2007). Although the national sample of SAT scores is more generalizable than Steele's and Aronson's, it also remains a selective sample that excludes individuals who did not take the SAT. There are a number of reasons why an

individual would not take the SAT (e.g., high school drop-out, no interest in college), and it is likely that different proportions of each racial group take the test. If a representative sample of SAT scores were available for the general population, it would likely be more discrepant from Steele's sample than the available national SAT scores of voluntary test-takers. Thus, Steele and Aronson (1995) investigated the effect of ST elicitors on highly-achieving Blacks who perform significantly better on standardized tests than either average Blacks or average Whites.

In Brown's and Day's (2006) study, the reported mean ACT score for Blacks was about 22.7. This was calculated assuming that the researchers obtained equal sample sizes of Blacks across levels of TD, which they did not. However, their cell sizes were close (17, 19, 17), and this is a reasonable estimate. A large-sample study reported the mean ACT score for Blacks as 17.1, with a standard deviation of 3.30 (Noble, 2003). Brown's and Day's (2006) sample of Blacks exhibited a mean ACT score that was 1.2 standard deviations above the national Black mean and was even 0.46 standard deviations above the national mean of Whites (see Noble, 2003). As discussed in the literature review (Chapter II) of this report, the mean performance of Blacks on the APM IQ test in Brown's and Day's sample across TD levels was about 102.7. This is well above the reported mean IQ score of Blacks (85) and is even above the overall mean, regardless of race (100). Thus, Brown and Day (2006) investigated the effect of ST elicitors on highly-achieving Blacks who perform significantly better on standardized tests than either average Blacks or average Whites.

The study conducted for this report analyzed a sample of Black participants with a mean ACT score of 16.3. This Black mean is only 0.24 standard deviations different from the national mean of 17.1. As previously mentioned, national norms for ACT/SAT scores are probably slightly higher than the mean of a sample that is representative of the full continuum of achievement and cognitive functioning. Thus, the slightly depressed mean of the Black sample in this study relative to the national mean suggests that it is reasonably representative of Blacks in the US and consequently, substantially more generalizable than previous ST studies. The mean ACT score for White participants in this study was 22.6. This White mean is only 0.43 standard deviations greater than the national mean (20.9). Table 11 presents the discrepancies in standard deviation units between ACT/SAT means of obtained samples in ST studies (including this study) and nationally representative means (nationally representative of test-takers, not of the entire population) for Blacks and Whites. These data illustrate that the samples obtained for this study are more generalizable than previous studies.

Table 11  
*Difference Between Means of Obtained Samples and National Scores by Study*

Study	Blacks	Whites
Steele and Aronson (1995; study 2)	+1.70	+1.27
Brown and Day (2006)	+1.20	+1.02
Current study	-0.24	+0.43

*Note.* Differences are reported in standard deviations

### New ST Effect

The results of this study showed that the influence of ST elicitors on a representative Black sample may not only be attenuated relative to high-achieving samples, but may actually reverse. Studies of the effect of ST elicitors on high achievers have shown that the presentation of ST elicitors adversely interfere with Blacks performance. The Black sample obtained for this study exhibited unimpaired performance when ST elicitors were presented relative to performance in a control condition which excluded all previously defined ST elicitors but maintained consequences for performance (e.g., scholarship, employment, school admission). The high-achieving Black samples in previous research showed elevated performance under a condition that removed all ST elicitors and presented no substantial performance consequence. The average-achieving Black sample obtained for this study showed depressed performance under this no stereotype threat condition. This study, when compared to previous research, suggests that the effect of the ST elicitors likely varies within racial groups depending on participants' typical performance on standardized tests. This finding suggests that much care must be given when generalizing the findings of ST research to specific populations.

### Uncompromised Psychometrics

Although differential performance on standardized tests is a problem, most interest in the issue stems from concern over differential performance on criteria which these standardized tests predict. Racial inequity persists across numerous



such correlates (predicted criteria): college performance, occupational performance, criminality and other maladaptive behaviors, welfare dependence, special education needs, drug use, income, etc. The ST model implicitly predicts that standardized tests will have limited predictive validity for Blacks due to the influence of ST elicitors. However, decades of research that show standardized tests (achievement and cognitive) have excellent validity, reliability, and factorial similarity when comparing the measures' utility for Blacks and Whites (see Herrnstein & Murray, 1994; Jensen, 1998). The ST model implicitly predicts that standardized tests would substantially underpredict Blacks' performance on related criteria. In fact, many such measures actually over-predict Black performance on many criteria (Herrnstein & Murray, 1994; Jensen, 1998; Levin, 1997; Noble, 2003). It is unclear whether ST theorists are cognizant that the likely goal of many consumers of racial differences research is to improve the situation for Blacks on the aforementioned correlated criteria (e.g., college performance) and not to compromise the predictive validity of the standardized tests.

## Conclusion

### New ST Phenomenon

Despite the evident invalidation of the ST theory by (1) previous racial differences research, (2) ST research itself, and/or (3) logical reasoning, popularity of the theory remains relatively unconstrained. Based on Psychinfo

searches (in January 2007), over 235 works have been published concerning stereotype threat since its 1995 inception. Unfortunately, some of those sources provide practitioners with methods of applying the theory (see Aronson, 2002; Aronson, Fried, & Good, 2002; Good, Aronson, & Inzlicht, 2003; Steele, Spencer, & Aronson, 2002). A Google internet search (January 2006) yielded 945,000 results. ST theory was discussed in an edition of the television program, *Frontline* (Chandler, 1999; see also Sackett, Hardison, & Cullen, 2004), in the television program 20/20 (ABC network, September 15, 2006), in the motion picture, *The Perfect Score* (Robbins, 2004), and in the bestselling book, *Blink* (Gladwell, 2005). Claude Steele has lectured on the theory at prestigious universities and has even given expert legal testimony concerning the theory's implications (Expert, n.d.). Millions of research dollars have been invested in ST research. Grants listed on Claude Steele's vitae alone sum to \$1,484,425 (Steele, n.d.). The theory is commonly covered in introductory psychology books, social psychology books, and sociology books. ST theory has propelled the careers of a number of researchers. Claude Steele was bestowed with the APA's 2003 Distinguished Scientific Contribution Award and the APA's 2003 Distinguished Contributions to Psychology in the Public Interest Award for his work in this area (Kersting, 2003).

This widespread popularity of stereotype threat theory is based largely on pervasive mischaracterization of the empirical evidence. As previously mentioned, ST theory is almost always mischaracterized as demonstrating equalization in racial groups' performance. Sackett and colleagues (2004a)

reported that 90.9% of scientific articles sampled characterized the stereotype threat incorrectly; 87.5% of popular media articles characterized the stereotype threat incorrectly; and 56% of applicable introductory psychology texts sampled characterized the stereotype threat incorrectly.

Because the inadequacies of stereotype threat theory are so numerous, the momentum that the theory has acquired is truly astonishing. The tendency for accomplished researchers to subscribe to the ST model despite its flaws is a phenomenon worthy of study and is perhaps the more appropriate referent for the term, *stereotype threat phenomenon*. It is not clear whether ST subscribers (1) lack competence to critically evaluate the ST research, (2) are disinterested in contemplating the validity of the ST interpretations, (3) or are knowingly endorsing a false theory. Nevertheless, the fact that so many academicians could perpetuate an elementary mistake in statistical interpretation exemplifies the need for more scrutiny and criticism, especially before applying theories in public service.

### Argumentum Ad Hominem

As discussed in the literature review, readers sometimes perceive attacks on non-hereditarian theories like ST as an endorsement of racist ideologies. Critics of ST theory are not ipso facto racists, and such prejudice is inhibiting scientific progress. It should be disconcerting that some academicians' egalitarian sensibilities may take precedence over their professional objectivity. The egalitarian struggle to ameliorate racial inequalities is noble, but it will not

succeed by championing fallacies. Continuing to research stereotype threat in relation to Black performance on standardized tests when a preponderance of research shows that stereotype threat elicitors cannot alleviate the gap is disingenuous to Blacks, usurps millions of research dollars which could be better allocated, and monopolizes other resources of the scientific community that could otherwise focus on searching for a solution to the racial gap.

This report criticizes of stereotype threat and acknowledges the work of disdained hereditarians, not to stifle egalitarian progress, but to advance it. Egalitarian psychologists who are interested in addressing the problem of depressed African American cognitive performance must not allow the divisive nature of the topic to deter them from their scientific objectivity. Just as Walker (2001) warned Afrocentrists that trying to create a more therapeutic version of history that never existed will inevitably injure the Black community, psychologists will not help the Black community by endorsing the feel-good, but debunked idea that the racial gap in standardized test performance can be explained by characteristics of the context in which the tests are administered. Attention should be redirected to the variety of factors (1) on which we can intervene, (2) that have been shown to affect test performance, (3) and by which racial inequity persists. Research has shown that the racial gap is likely influenced by a complex constellation of such variables, which include but are certainly not limited to: cultural values, poverty, teacher prejudice, deficient parental involvement, inadequate schools, educator prejudice, insufficient healthcare, drug use, exposure to toxins during development, psychological

maladjustment, learned helplessness, educational apathy/ anti-intellectualism, lack of learning opportunity, negative peer pressure, and lack of academic role models.

### Reasonable Suspicion

This report reviewed a variety of contextual issues, empirical trends, and alternative explanations to provide a more sophisticated understanding of stereotype threat theory. Stereotype threat theory exists in the larger context of racial differences research which comprises numerous factions of scholars and opposing worldviews. The biased context in which ST theory was formulated, investigated, published, and applied is a critical component to understanding its interpretation by others. As explicitly discussed in many sources (see Browne-Miller, 1995; Herrnstein & Murray, 1994; Levin, 1997; Tucker, 1994), racial differences research has extensive implications across a broad range of issues including, but certainly not limited to: affirmative action; employment; immigration; welfare; capital punishment and other penal legislation; mental health services; special education; and compensatory education. Racial differences research is of profound importance to school psychologists who must routinely assess the cognitive abilities of Black children. To subscribe to ST Theory without subjecting it to reasonable criticism is to inhibit progress in alleviating racial inequality. Most importantly, the study of the mischaracterization, momentum, and immunity of ST theory to criticism serves as an ominous reminder to the scientific community that subjectivity continues to have a profound influence on research. The lack of

criticism of ST theory is disheartening. Social and scientific progress cannot advance without reason, and reason cannot exist without criticism.

“Reason must in all its undertakings subject itself to criticism, should it limit freedom of criticism by any prohibitions it must harm itself, drawing upon itself a damaging suspicion. Nothing is so important through its usefulness, nothing so sacred, that it may be exempted from this searching examination, which knows no respect for persons. Reason depends on this freedom for its very existence.”

(Kant, 1781, p593)

## REFERENCES

- Allport, G. (1954). *The nature of prejudice*. New York: Addison-Wesley.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed. text revision). Washington DC: Author.
- Aronson, J. (Ed.).(2002). *Improving academic achievement: Impact of psychological factors on education*. New York: Academic Press.
- Aronson, J., Fried, C.B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125.
- Aronson, J. Lustina, M.J., & Good, C. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 1999.
- Artiles, A.J., & Trent, S.C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education, 27* (4), 410-37.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28* (2), 117-148.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: W.H. Freeman.
- Beck, F.W. (1984). Another look at the effects of the System of Multicultural Pluralistic Assessment. *Journal of School Psychology, 22* (4), 347-52.
- Blascovich, J., Spencer, S.J., Quinn, D., & Steele, C.M. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*(3), 225-229.
- Bouchard, T. J. Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science, 250*, 223-228.
- Braden, J.P. (1995). For whom "the bell" tolls: Why *The Bell Curve* is important for school psychologists. *School Psychology Review, 24* (1), 27-35.



- Brand, C. (1996). *The g factor*. Wiley & Sons Ltd.:UK. (revised edition available for download at <http://www.douance.org/qi/brandtgf.htm>).
- Bronfenbrenner, U. (1993). The ecology of cognitive development: Research models and fugitive findings. In R.H. Wozniak & K.W. Fischer (Eds.), *Development in action: Acting and thinking in specific environments* (3-44). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, J.L. (2001). Performance expectations are not a necessary mediator of stereotype threat in African American verbal test performance. *Dissertation Abstracts International*, 61(11), 6184B. (obtained from Digital Dissertations).
- Brown, R.P., Day E.A. (2006) The difference isn't black and white: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology*, 91(4), 979-985.
- Browne-Miller, A. (1995) *Intelligence policy : its impact on college admissions and other social policies*. New York: Plenum Press.
- Cadinu, M., Maass, A., & Frigerio, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, 33(2), 267-285.
- Chabris, C. (1998). IQ since *The Bell Curve*. *Commentary* (August, 1998).
- Chandler, M. (Writer & Director). (1999, October 4). Secrets of the SAT [Television series episode]. In M. Sullivan (Executive Producer), *Frontline*. Boston: WGBH.
- Chapman, M. (Ed.). (1993). *Social and biological aspects of ethnicity*. Oxford: Oxford University Press.
- Clark, K.B. (1955). *Prejudice and your child*. Boston: Beacon Press.
- Cokley, K.O. (2002). Ethnicity, gender and academic self-concept: A preliminary examination of academic disidentification and implications for psychologists. *Cultural Diversity & Ethnic Minority Psychology*, 8(4), 378-388.
- College Board. (1996). *1996 National Report: 1996 Background Information*. Retrieved January 19, 2007 from [http://www.collegeboard.com/about/news\\_info/cbsenior/yr1996/nat/cbtbk296.html](http://www.collegeboard.com/about/news_info/cbsenior/yr1996/nat/cbtbk296.html)

- Croizet, J-C., Desert, M., & Dutrevis, M. (2001). Stereotype threat, social class, gender, and academic under-achievement: When our reputation catches up to us and takes over. *Social Psychology of Education*, 4(3-4), 295-310.
- Cullen, M.J., Hardinson, C.M., & Sackett, P.R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, 89 (2), 220-230.
- Davis III, C. (2000). Identity in context: The relationship between Black racial identity and stereotype threat. *Dissertation Abstracts International*, 61(1), 586B. (obtained from Digital Dissertations).
- Desert, M., Croizet, J-C., & Leyens, J-P. (2002). Stereotype threat: An interaction between situation and identity. *Annee Psychologique*, 102(3), 555-576.
- Devine, P.G., & Elliot, A.J.(1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality & Social Psychology Bulletin*, 21(11), 1139-1150.
- Devlin, B., Feinberg, S.E., Resnick, D.P., & Roeder, K. (Eds.). (1997). *Intelligence, genes, and success: Scientists respond to The Bell Curve*. New York: Copernicus.
- Dorans, N.J. (1999). *Correspondences Between ACT and SAT I Scores*. Retrieved January 19, 2007 from [http://www.collegeboard.com/research/pdf/rr9901\\_3913.pdf](http://www.collegeboard.com/research/pdf/rr9901_3913.pdf)
- Expert Testimony of Claude Steele*. (n.d.). Retrieved April 9, 2005 from <http://www.umich.edu/~urel/admissions/legal/expert/steele.html>.
- Fischer, C.S., Hout, M., Jankowski, M.S., Lucas, S.R., Swindler, A., & Voss, K. (1996). *Inequality by design: Cracking The Bell Curve myth*. Princeton, NJ: Princeton University Press.
- Flynn, J.R. (2003). Movies about intelligence: The limitations of g. *Current Directions in Psychological Science*, 12 (3), 95-99.
- Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the "burden of acting White. *Urban Review*, 18(3), 176-206.
- Fraser, S. (1995). *The Bell Curve wars: Race, intelligence, and the future of America*. New York: BasicBooks.
- Frisby, C.L. (1993a). "Afrocentric" explanations for school failure: symptoms of denial, frustration, and despair. *School Psychology Review*, 22 (3), 568-578.

- Frisby, C.L. (1993b). One giant step backward: Myths of Black cultural learning styles. *School Psychology Review*, 22 (3), 535-558.
- Frisby, C.L. (1995). When facts and orthodoxy collide: The bell curve and the robustness criterion. *School Psychology Review*, 24 (1), 12-19.
- Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. 1892 ed. London: Macmillan.
- Gladwell, M. (2005). *Blink*. New York: Little, Brown, and Company.
- Goffman, I. (1963). *Stigma*. New York: Simon & Shuster, Inc.
- Goldhaber, D.E. (2000). *Theories of human development: Integrative perspectives*. Mountain View, CA: Mayfield Publishing Company.
- Gonzales, P.M., Blanton, H., Williams, K.J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality & Social Psychology Bulletin*, 28(5), 659-670.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology*, 24, 645-662.
- Goodman, J.F. (1977). The diagnostic fallacy: A critique of Jane Mercer's concept of mental retardation. *Journal of School Psychology*, 15, 198-205.
- Gottfredson, L.S., (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography, *Intelligence*, 22(1), 13-23.
- Gould, S.J. (1996). *The mismeasure of man*. New York: W.W. Norton.
- Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47(9), 1083-1101.
- Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Simon and Schuster.
- Hickman, J.A., & Reynolds, C.R., (1986). Are race differences in mental test scores an artifact of psychometric methods? A test of Harrington's experimental model. *The Journal of Special Education*, 20 (4), 409-430.
- Howe, S. (1998). *Afrocentrism*. New York: Verso

- Jacoby, R., & Glauberman, N. (Eds.). (1995). *The Bell Curve debate: History, documents, opinions*. New York: Random House.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.
- Jensen, A.R. (1984). The Black-White difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18* (3), 377-408.
- Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT.: Praeger.
- Jirsa, J.E. (1983). The SOMPA: A brief examination of technical considerations, philosophical rationale, and implications for practice. *Journal of School Psychology, 21* (1), 13-21.
- Josephs, R.A., Newman, M.L., & Brown, R.P. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science, 14*(2), 158-163.
- Kant, I. (1781). *Critique of Pure Reason*. Retrieved January 17, 2007 from <http://www.hkbu.edu.hk/~ppp/cpr/dmeth.html>
- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality & Social Psychology Bulletin, 29*(3), 371-381.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. New Jersey: Prentice-Hall, Inc.
- Kersting, K. (2003). Countering insidious stereotypes: Hidden threats to minority groups can be overcome through exposure. *Monitor On Psychology, 34* (9), 34-35.
- Kincheloe, J.L., Steinberg, S.R., & Gresson III, A.D. (Eds.).(1996). *Measured lies: The Bell Curve examined*. New York: St. Martin's Press.
- Kush, J.C., Watkins, M.W., Ward, T.J., Sandra, B., Canivez, G.L., & Worrell, F.C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review, 30* (1), 70-89.
- Levin, M.E.(1997). *Why race matters: race differences and what they mean*. Westport, CT: Praeger.
- Levin, M.E.(1995). Does race matter? *American Psychologist, 50*(1), 45-46.

- Lieberman, L. (2001). How "Caucasoids" got such big crania and why they shrank: From Morton to Rushton. *Current Anthropology*, 42(1), 69-95.
- Loveless, Peggy (2000). Does self-stereotyping affect performance on the ACT assessment? *Research Report Series*. Retrieved December 28, 2006 from [http://www.act.org/research/reports/pdf/ACT\\_RR2000-6.pdf](http://www.act.org/research/reports/pdf/ACT_RR2000-6.pdf)
- Marable, M. (Ed.). (2000). *Dispatches from the ebony tower: Intellectuals confront the African American experience*. New York: Colombia University Press.
- Mash, E.J., & Barkley, R.A. (Eds.). (1996). *Child psychopathology*. New York: The Guilford Press.
- Mayfield, J.W., & Reynolds, C.R. (1997). Are ethnic differences in diagnosis of childhood psychopathology an artifact of psychometric methods? An experimental evaluation of Harrington's hypothesis using parent reported symptomology. *Journal of School Psychology*, 36 (3), 313-334.
- McDermott, P.A., & Spencer, M.B. (1997). Racial and social class prevalence of psychopathology among school-age youth in the U.S. *Youth & Society*, 28 (4), 387-414.
- McGrew, K.S., & Flanagan, D.P. (1998). *The intelligence test desk reference: Gf-Gc cross battery assessment*. Boston: Allyn and Bacon.
- McKay, P.F. (1999). Stereotype threat and its effect on the cognitive ability test performance of African-Americans: The development of a theoretical model. *Dissertation Abstracts International*, 60(4), 1897B. (obtained from Digital Dissertations).
- McKay, P.F., Doverspike, D., Bowen-Hilton, D., & Martin, Q.D. ( 2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African-Americans. *Journal of Applied Social Psychology*, 32(4), 767-787.
- McKay, P.F., Doverspike, D., Bowen-Hilton, D., & McKay, Q.D. (2003). The effects of demographic variables and stereotype threat on Black/White differences in cognitive ability test performance. *Journal of Business and Psychology*, 18 (1), 1-13.
- Miele, F. (2002). *Intelligence, race, and genetics: conversations with Arthur R. Jensen*. Boulder, CO: Westview Press.

- Molnar, S. (1983). *Human variation: races, types, and ethnic groups*. Englewood Cliffs, NJ: Prentice-Hall.
- Neisser, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., Sternberg, R., & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, 51(2), 77-101.
- Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington, DC: APA
- Noble, J.P., Roberts, W.L., & Sawyer R.L. (2006). Student achievement, behavior, perceptions, and other factors affecting ACT scores. *ACT Research Report Series*. Retrieved December 28, 2006 from [http://www.act.org/research/reports/pdf/ACT\\_RR2006-1.pdf](http://www.act.org/research/reports/pdf/ACT_RR2006-1.pdf)
- Noble, J. (2003). The effects of using ACT composite score and high school average on college admission decisions for racial/ethnic groups. *ACT Research Report Series*. Retrieved December 28, 2006 from [http://www.act.org/research/reports/pdf/ACT\\_RR2003-1.pdf](http://www.act.org/research/reports/pdf/ACT_RR2003-1.pdf)
- Novak, M. (1994). Sins of the cognitively elite. *National Review*, Dec 5, p 58-61.
- Oakland, T. (1995). The bell curve: Some implications for the discipline of school psychology. *School Psychology Review*, 24 (1), 20-26.
- O'Brien, L..T., & Crandall, C.S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality & Social Psychology Bulletin*, 29(6), 782-789.
- Ogbu, J.U. (1986). The consequences of the American caste system. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives* (19-56). Hillsdale, NJ: Erlbaum.
- Ogbu, J. U. (1978). *Minority education and caste: The American system in cross-cultural perspective*. New York: Academic Press.
- Ogbu, J. U. (1994). From cultural differences to differences in cultural frames of reference. In P. M. Greenfield & R. R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 365-391). Hillsdale, NJ: Erlbaum.
- Oswald, D.P., Coutinho, M.J., Best, A.M., & Singh, N.N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *Journal of Special Education*, 32(4), 194-206.

- Overfield, T. (1995). *Biological variations in health and illness: race, age, and sex differences*. New York: CRC Press.
- Raven, J., Raven, J.C., & Court, J.H. (1998a). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1 general overview*. Oxford: OPP Limited.
- Raven, J., Raven, J.C., & Court, J.H. (1998b). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 4 general overview*. Oxford: OPP Limited.
- Robinson-Heath, D.C. (2002). Achievement motivation: A comparison of African-American achieving and underachieving students. *Dissertation Abstracts International*, 62(8), 2683A. (obtained from Digital Dissertations).
- Robbins, B. (Director). (2004). *The Perfect Score* [motion picture]. Hollywood: Paramount Pictures.
- Rushton, J.P., & Rushton, E.W. (2003). Brain size, IQ, and racial-group differences: Evidence from musculoskeletal traits. *Intelligence*, 31(2), 139-155.
- Sackett, P.R., Schmitt, N., Ellingson, J.E., Kabin, M.B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist*, 56 (4), 302-318.
- Sackett, P.R., Hardison, C.M., & Cullen, M.J. (2004a). On interpreting stereotype threat as accounting for African-American differences on cognitive tests. *American Psychologist*, 59 (1), 7-13.
- Sackett, P.R., Hardison, C.M., & Cullen, M.J. (2004b). On the Value of Correcting Mischaracterizations of stereotype threat Research. *American Psychologist*, 59 (1), 48-49.
- Sattler, J.M. (2001). *Assessment of children: Cognitive applications*. San Diego: J.M. Sattler, Publisher, INC.
- Scarr, S. (1998). On Arthur Jensen's integrity. *Intelligence*, 26(3), 227-232.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality & Social Psychology*, 85(3), 440-452.
- Smith, J.L., & White, P.H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, 47(3-4), 179-191.

- Smith, J.L. (2004). Understanding the Process of stereotype threat: A Review of *Mediational Variables and New Performance Goal Directions*. *Educational Psychology Review*, 16(3), 177-206.
- Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42(2), 137-144.
- Spencer, S.J., Steele, C.M., & Quinn, D.M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4-28.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality & Social Psychology*, 75(5), 1191-1197.
- Steele, C. M. (n.d.) *Curriculum Vitae*. Retrieved April 9, 2005 from <http://www-psych.stanford.edu/~steele/vita.html>.
- Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52 (6), 613-629.
- Steele, C.M. & Aronson, J. (1995). Stereotype threat and intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69 (5), 797-811.
- Steele, C. M., & Aronson, J. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist*, 59 (1), 47-48.
- Steele, C.M., Spencer, S.J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In Zanna, M.P. (Ed), *Advances in experimental social psychology* (v 34, pp 379-440). San Diego, CA: Academic Press, Inc.
- Sternberg, R.J., Grigorenko, E.L., Kidd, K.K. (2005). Intelligence, race, and genetics. *American Psychologist*, 60(1), 46-59.
- Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality & Social Psychology Bulletin*, 28(12), 1667-1678.
- Stone, J., Lynch, C.I., & Sjomeling, M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality & Social Psychology*, 77(6), 1213-1227.



- Tucker, W.H. (1994). *The science and politics of racial research*. Chicago: University of Illinois Press.
- Valencia, R.R., & Suzuki, L.A. (2001). *Intelligence testing and minority students : foundations, performance factors, and assessment issues*. London: Sage Publications.
- Viney, W., & King, D.B. (2003). *A history of psychology: Ideas and context*. 3<sup>rd</sup> edition. Allyn and Bacon: Boston.
- Walker, C.E. (2001). *We can't go home again: An argument about afrocentrism*. New York: Oxford University Press.
- Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*(3-4), 219-240.
- Whaley, A.L. (1998). Issues of validity in empirical tests of stereotype threat theory. *American Psychologist, 53*(6), 679-680.
- Wicherts, J.M. (2005). Stereotype threat research and the assumptions underlying analysis of covariance. *American Psychologist, 60*(3), 267-269.
- Weiss, M.L., & Mann, A.E. (1981). *Human biology and behavior*. Boston: Little, Brown, and Company.
- Wurtz, R.G., Sewell, T., & Manni, J.L. (1985). The relationship of Estimated Learning Potential to performance on a learning task and achievement. *Psychology in the Schools, 22* (3), 293-302.


## APPENDICES

### Appendix A *Cognitive Task Cover for the Control Condition*

1

<p><b>INTEGRATED SCHOLASTIC - OCCUPATIONAL TEST®</b></p>	<p><i>Institutional Version for use in conjunction with research.</i></p> <p>ASSOCIATED STUDY TITLE: ISOT Compensatory Provision for Oklahoma Partners and Group Data Collection</p>
--	--

ISOT and ETS are registered trademarks of Educated Testing Services. © 2005. All rights reserved.  
Dissemination of test materials strictly prohibited. www.ets.org



Do not mark in this book.                      Do not open until instructed to do so.

Appendix B

*Cognitive Task Cover for the Stereotype Threat Condition*

1

Diagnostic Test of Verbal Ability and Intelligence	STUDY TITLE: Understanding Personal Factors that Affect Verbal Ability and Intelligence
---	---

\* Please be sure to accurately record your race on the scoring protocol.

Do not mark in this book.

Do not open until instructed to do so.

Appendix C

*Cognitive Task Cover for the No Stereotype Threat Condition*

Problem Solving Challenges	STUDY TITLE: Understanding Different Methods for Solving Problems
-------------------------------	--

Do not mark in this book.

Do not open until instructed to do so.

Appendix D  
*OSU IRB approval letter*

**Oklahoma State University Institutional Review Board**

Date: Tuesday, January 10, 2006  
IRB Application No ED0663  
Proposal Title: Stereotype Threat Theory as an Explanation for the Depressed Performance on Cognitive Ability Measures by African Americans

Reviewed and Processed as: Expedited

**Status Recommended by Reviewer(s): Approved Protocol Expires: 1/9/2007**

Principal Investigator(s)

John M. Nomura 314 N. Wildwood Acres Ct Stillwater, OK 74075	Terry Stinnett 434 Willard Stillwater, OK 74078
--	---

---

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 415 Whitehurst (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,



Sue C. Jacobs, Chair  
Institutional Review Board

Appendix E  
Langston University IRB approval letter

**Langston University  
Institutional Review Board  
Human Subjects Review**

Date February 10, 2006

IRB # 42

Proposal Title: Stereotype Threat Theory as an Explanation for the Depressed Performance on Cognitive Ability Measures by African Americans

Principal Investigator (s): Mr. John Nomura  
Reviewed and Processed: February 7, 2006

IRB Decision Rendered: Approved  
(approval or disapproval)

All approvals must be subject to review by a full Institutional Review Board at the next meeting as well as subject to the monitoring process of the Board at any time during the approval period.

Approval status period valid for data collection is one calendar year. A request for continuation of a research project beyond the one-year time must be submitted to the Board in writing prior to the one-year expiration date.

Any changes or modifications to the approved project must also be submitted for approval.

Comments: **Langston University contact information must be added to informed consent form.**

Modifications or Terms and Conditions for Approval:

Reason(s) for Disapproval: \_\_\_\_\_

Signature *Yvonne Montgomery* Date 2/10/06  
Chair of Institutional Review Board

Appendix F  
OSU University Registrar Approval Letter



Office of the Registrar  
322 Student Union  
Stillwater, Oklahoma 74078-1013  
Fax: 405-744-8426  
<http://www.okstate.edu/registrar>

Certifications: 405-744-6873  
Fee Adjustments: 405-744-7663  
Transcripts: 405-744-6888  
Veterans' Service: 405-744-6868  
Registration Services: 405-744-6890  
General Information: 405-744-6876

August 29, 2005

To Whom It May Concern:

This letter is to inform you that I have read the proposal for John Nomura's research project with the School of Psychology. His project is titled "Stereotype Threat Theory as an Explanation for Depressed Cognitive Performance of African Americans".

The Registrar's Office has agreed to provide relevant data for this research project. Each individual participant in the project will be required to sign a consent form that will be submitted to the Registrar's Office. It is the understanding of the Registrar's Office that the requested data will include ACT and SAT scores of the individual participants. All researchers involved in the study will be required to sign a confidentiality form approved by the IT Security Department. The researchers have agreed to use random numbers in lieu of identifying information in their working files. Researchers have further agreed to restrict identifying information to a single encrypted file and not keep that file on a personal computer.

If you have any questions concerning this matter, please do not hesitate to contact me.

Sincerely,

A handwritten signature in cursive script that reads "Joan Payne".

Joan Payne  
Associate Registrar

Appendix G  
Langston University Registrar Approval Letter



**LANGSTON  
UNIVERSITY**

Office of Enrollment Management

*Registrar • Records • Registration • Admissions*

September 30, 2005

Mr. John Nomura  
314 N. Wildwood Acres CT  
Stillwater, OK 74075

To Whom It May Concern:

REF: OSU RESEARCH

The purpose of this letter is to verify the Registrar's Office cooperation with the Oklahoma State University School of Psychology Department's research project currently titled: "Stereotype Threat Theory as an Explanation for the Depressed Cognitive Performance of African Americans."

If written permission is obtained from each student participant and given to the Registrar's Office at Langston University, we will provide the relevant data, with the understanding this data will include ACT and SAT scores on individual research participants. We stipulate that the researchers must use methods of data security approved by us and must maintain strict confidentiality of records. Researchers have agreed to use random numbers in lieu of identifying information in their working files. Researchers have further agreed to restrict identifying information to a single encrypted file and not keep that file on a personal computer.

Sincerely,

Margie Allen Bonner  
Registrar

*An Equal Opportunity/Affirmative Action Employer*

# Appendix H

## Response Sheet Front for the Control Condition

LAST NAME (space) FIRST NAME (space) M.I.															SOCIAL SECURITY NUMBER			
															Last 4 digits only			
(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(0)	(0)	(0)	(0)
(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(B)	(1)	(1)	(1)	(1)
(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(2)	(2)	(2)	(2)
(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(3)	(3)	(3)	(3)
(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(4)	(4)	(4)	(4)
(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(F)	(5)	(5)	(5)	(5)
(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(G)	(6)	(6)	(6)	(6)
(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(7)	(7)	(7)	(7)
(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(I)	(8)	(8)	(8)	(8)
(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(J)	(9)	(9)	(9)	(9)
(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(K)	(0)	(0)	(0)	(0)
(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(L)	(1)	(1)	(1)	(1)
(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(M)	(2)	(2)	(2)	(2)
(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(3)	(3)	(3)	(3)
(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(4)	(4)	(4)	(4)
(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(P)	(5)	(5)	(5)	(5)
(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(Q)	(6)	(6)	(6)	(6)
(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(7)	(7)	(7)	(7)
(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(S)	(8)	(8)	(8)	(8)
(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(T)	(9)	(9)	(9)	(9)
(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(U)	(0)	(0)	(0)	(0)
(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(V)	(1)	(1)	(1)	(1)
(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(W)	(2)	(2)	(2)	(2)
(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(3)	(3)	(3)	(3)
(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(Y)	(4)	(4)	(4)	(4)
(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(5)	(5)	(5)	(5)

SITE CODE	BOOK CODE
(0)	(0)
(1)	(1)
(2)	(2)
(3)	(3)
(4)	(4)
(5)	(5)
(6)	(6)
(7)	(7)
(8)	(8)
(9)	(9)

**INTEGRATED SCHOLASTIC - OCCUPATIONAL TEST**

I.S.O.T. and ETS are registered trademarks of Educated Testing Services. Copyright 2005. All rights reserved. Dissemination of test materials strictly prohibited. www.ets.org

2006 edition. Form E2. N.America.

**WRITE:** I hereby certify that I am the person signing below.

---



---



---



---

**SIGN YOUR NAME:**

---

# Appendix I

## Response Sheet Back for the Control Condition

Fill in circle completely. Make all erasures complete. Use No. 2 pencil only.

1 (A) (B) (C) (D) (E)	17 (1) (2) (3) (4) (5)	33 (1) (2) (3) (4) (5) (6) (7) (8)	41 (1) (2) (3) (4) (5) (6) (7) (8)	49 (1) (2) (3) (4) (5) (6) (7) (8)	57 (1) (2) (3) (4) (5) (6) (7) (8)
2 (1) (2) (3) (4) (5)	18 (1) (2) (3) (4) (5)	34 (1) (2) (3) (4) (5) (6) (7) (8)	42 (1) (2) (3) (4) (5) (6) (7) (8)	50 (1) (2) (3) (4) (5) (6) (7) (8)	58 (1) (2) (3) (4) (5) (6) (7) (8)
3 (1) (2) (3) (4) (5)	19 (1) (2) (3) (4) (5)	35 (1) (2) (3) (4) (5) (6) (7) (8)	43 (1) (2) (3) (4) (5) (6) (7) (8)	51 (1) (2) (3) (4) (5) (6) (7) (8)	59 (1) (2) (3) (4) (5) (6) (7) (8)
4 (1) (2) (3) (4) (5)	20 (1) (2) (3) (4) (5)	36 (1) (2) (3) (4) (5) (6) (7) (8)	44 (1) (2) (3) (4) (5) (6) (7) (8)	52 (1) (2) (3) (4) (5) (6) (7) (8)	60 (1) (2) (3) (4) (5) (6) (7) (8)
5 (1) (2) (3) (4) (5)	21 (1) (2) (3) (4) (5)	37 (1) (2) (3) (4) (5) (6) (7) (8)	45 (1) (2) (3) (4) (5) (6) (7) (8)	53 (1) (2) (3) (4) (5) (6) (7) (8)	61 (1) (2) (3) (4) (5) (6) (7) (8)
6 (1) (2) (3) (4) (5)	22 (1) (2) (3) (4) (5)	38 (1) (2) (3) (4) (5) (6) (7) (8)	46 (1) (2) (3) (4) (5) (6) (7) (8)	54 (1) (2) (3) (4) (5) (6) (7) (8)	62 (1) (2) (3) (4) (5) (6) (7) (8)
7 (1) (2) (3) (4) (5)	23 (1) (2) (3) (4) (5)	39 (1) (2) (3) (4) (5) (6) (7) (8)	47 (1) (2) (3) (4) (5) (6) (7) (8)	55 (1) (2) (3) (4) (5) (6) (7) (8)	
8 (1) (2) (3) (4) (5)	24 (1) (2) (3) (4) (5)	40 (1) (2) (3) (4) (5) (6) (7) (8)	48 (1) (2) (3) (4) (5) (6) (7) (8)	56 (1) (2) (3) (4) (5) (6) (7) (8)	
9 (1) (2) (3) (4) (5)	25 (1) (2) (3) (4) (5)				
10 (1) (2) (3) (4) (5)	26 (1) (2) (3) (4) (5)				
11 (1) (2) (3) (4) (5)	27 (1) (2) (3) (4) (5)				
12 (1) (2) (3) (4) (5)	28 (1) (2) (3) (4) (5)				
13 (1) (2) (3) (4) (5)	29 (1) (2) (3) (4) (5)				
14 (1) (2) (3) (4) (5)	30 (1) (2) (3) (4) (5)				
15 (1) (2) (3) (4) (5)	31 (1) (2) (3) (4) (5)				
16 (1) (2) (3) (4) (5)	32 (1) (2) (3) (4) (5)				

**I.S.O.T.**



Appendix J  
Response Sheet for the Stereotype Threat Condition

## Diagnostic Test of Verbal Ability and Intelligence

STUDY TITLE: Understanding Personal Factors that Affect Verbal Ability and Intelligence

PLEASE PRINT CLEARLY.

Name: \_\_\_\_\_

Last 4 digits of your SSN:

\_\_\_\_ \_

What is your race?

\_\_\_\_\_

- |          |          |          |
|----------|----------|----------|
| 1. ____  | 13. ____ | 25. ____ |
| 2. ____  | 14. ____ | 26. ____ |
| 3. ____  | 15. ____ | 27. ____ |
| 4. ____  | 16. ____ | 28. ____ |
| 5. ____  | 17. ____ | 29. ____ |
| 6. ____  | 18. ____ | 30. ____ |
| 7. ____  | 19. ____ | 31. ____ |
| 8. ____  | 20. ____ | 32. ____ |
| 9. ____  | 21. ____ |          |
| 10. ____ | 22. ____ |          |
| 11. ____ | 23. ____ |          |
| 12. ____ | 24. ____ |          |

- |          |          |          |
|----------|----------|----------|
| 33. ____ | 45. ____ | 57. ____ |
| 34. ____ | 46. ____ | 58. ____ |
| 35. ____ | 47. ____ | 59. ____ |
| 36. ____ | 48. ____ | 60. ____ |
| 37. ____ | 49. ____ | 61. ____ |
| 38. ____ | 50. ____ | 62. ____ |
| 39. ____ | 51. ____ |          |
| 40. ____ | 52. ____ |          |
| 41. ____ | 53. ____ |          |
| 42. ____ | 54. ____ |          |
| 43. ____ | 55. ____ |          |
| 44. ____ | 56. ____ |          |

AO-1

Appendix K  
Response Sheet for the No Stereotype Threat Condition

## Problem Solving Challenges

STUDY TITLE: Understanding Different Methods for Solving Problems

PLEASE PRINT CLEARLY.

Name: \_\_\_\_\_

Last 4 digits of your SSN:

\_\_\_\_ \_

- |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|
| 1. ____  | 13. ____ | 25. ____ | 33. ____ | 45. ____ | 57. ____ |
| 2. ____  | 14. ____ | 26. ____ | 34. ____ | 46. ____ | 58. ____ |
| 3. ____  | 15. ____ | 27. ____ | 35. ____ | 47. ____ | 59. ____ |
| 4. ____  | 16. ____ | 28. ____ | 36. ____ | 48. ____ | 60. ____ |
| 5. ____  | 17. ____ | 29. ____ | 37. ____ | 49. ____ | 61. ____ |
| 6. ____  | 18. ____ | 30. ____ | 38. ____ | 50. ____ | 62. ____ |
| 7. ____  | 19. ____ | 31. ____ | 39. ____ | 51. ____ |          |
| 8. ____  | 20. ____ | 32. ____ | 40. ____ | 52. ____ |          |
| 9. ____  | 21. ____ |          | 41. ____ | 53. ____ |          |
| 10. ____ | 22. ____ |          | 42. ____ | 54. ____ |          |
| 11. ____ | 23. ____ |          | 43. ____ | 55. ____ |          |
| 12. ____ | 24. ____ |          | 44. ____ | 56. ____ |          |

1B

## VITA

John Michael Nomura

Candidate for the degree of

Doctor of Philosophy

Report: STEREOTYPE THREAT THEORY AS AN EXPLANATION FOR THE  
DEPRESSED PERFORMANCE ON COGNITIVE ABILITY MEASURES BY  
AFRICAN AMERICANS

Major Field: School Psychology

### Biographical:

Education: Graduated from Owasso High School, Owasso, Oklahoma in May 1994. Received Bachelor of Arts degree in Psychology from University of Oklahoma in December 2000. Completed the requirements for Masters of Science degree with a major in Educational Psychology at Oklahoma State University in July 2004.

Experience: Sand Springs Home for Children, Head Lifeguard and Activity Supervisor, 1993-1995; Griffin Memorial Hospital , Mental Health Aid, August 1996- May 1998; Hillcrest Positive Outcomes, Mental Health Tech, January 1999- June 1999; Griffin Memorial Hospital, Practicum Student, January 1999- May 1999; Oklahoma Youth Center , Patient Care Assistant, June 1999- July 2002; Richmond Elementary School, Practicum Student, August 2002- December 2002; Perry School District, Practicum Student, August 2003-May 2004; Weiss Psychology Lab , University of Oklahoma, Lead Research Assistant, July 1997-May 1999; Oklahoma State University, Department of School Psychology, Research Team Member, October 2002-May 2006; Oklahoma State University, Course Instructor, August 2002-July 2006; Cypress Fairbanks ISD APPIC Pre-doctoral Internship, intern, August 2006-June 2007.

Professional Memberships: American Psychological Association, Division 16; National Association of School Psychologists; Oklahoma School Psychological Association.

Name: John Nomura

Date of Degree: July 2007

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: STEREOTYPE THREAT THEORY AS AN EXPLANATION FOR  
THE DEPRESSED PERFORMANCE ON COGNITIVE ABILITY  
MEASURES BY AFRICAN AMERICANS

Pages in Study: 149

Candidate for the Degree of Doctor of Philosophy

Major Field: School Psychology

The purpose of this study was to investigate the effect of stereotype threat elicitors on the cognitive performance of African-Americans and to examine whether the stereotype threat effect can account for the discrepancy in performance between African-American students and Caucasian-American students on standardized cognitive tests. Unlike many previous studies, this study used no covariate for the primary analysis and employed a control condition that simulated a standardized testing environment. Participants were 60 Black and 60 White college students. The report also presents a balanced and critical review of the literature, along with a discussion on sources of bias. Results indicate that stereotype threat theory cannot account for the racial gap in cognitive performance. It is suggested that the interpretations of previous studies which support stereotype threat theory are predicated on statistical misuse, unrepresentative samples, and disregard for previous research on racial differences. Data were analyzed using ANOVA and ANCOVA to demonstrate the effect of the covariate.

ADVISER'S APPROVAL: Terry Stinnett