2007

Cognitive Ability, Job Knowledge, and Stereotype Threat: When does Adverse Impact Result?

Mark V. Palumbo
Wright State University

COGNITIVE ABILITY, JOB KNOWLEDGE, AND STEREOTYPE THREAT:

WHEN DOES ADVERSE IMPACT RESULT?

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

By

MARK V. PALUMBO
B.S., The Pennsylvania State University, 2000,
M.S., Wright State University, 2004

2007
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

August 6, 2007

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Mark V. Palumbo ENTITLED Cognitive Ability, Job Knowledge, and Stereotype Threat: When Does Adverse Impact Occur? BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Debra Steele-Johnson, Ph.D.
Dissertation Co-Director

Valerie L. Shalin, Ph.D.
Dissertation Co-Director

John M. Flach, Ph.D.
Department Chair

Joseph F. Thomas, Ph.D.
Dean, School of Graduate Studies

Committee on
Final Examination

Debra Steele-Johnson, Ph.D.

Valerie L. Shalin, Ph.D.

Corey E. Miller, Ph.D.

David M. LaHuis, Ph.D.

Martin P. Gooden, Ph.D.

ABSTRACT

Palumbo, Mark V. Ph.D., Department of Psychology, Wright State University, 2007. Cognitive Ability, Job Knowledge, and Stereotype Threat: When Does Adverse Impact Occur?

This research compared the efficacy of a cognitive ability test and two types of job knowledge tests for predicting job performance. Further, I examined job knowledge as a mechanism through which cognitive ability affects performance. Finally, I examined both types of tests relative to specific propositions from stereotype threat theory. Specifically, I examined the propositions that perceptions of the tests may cause mean score differences between Blacks and Whites and compared the effects of test perceptions relative to both test types.

Results demonstrated that job knowledge accounted for significantly more variance in task performance than cognitive ability. Furthermore, job knowledge completely mediated the effects of cognitive ability on performance. However, stereotype threat theory's proposed test perceptions failed to account for mean test score differences between the two groups. Rather, Blacks' *misperceptions* relative to what each test was designed to measure was found to be detrimental for test performance. Also, regardless of what the test was designed to measure, Blacks still perceived both types of tests as (stereotype) threatening.

TABLE OF CONTENTS

Page

Page

Page

Page

Page

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Initially, I would like to thank the co-chairpersons of my dissertation committee. I would like to thank Debra Steele-Johnson for her tireless commitment to the quality of this research. Her unwavering research ethic has allowed me to strive for excellence in my work and has defined my own level of expectation for all future endeavors. Thank you, Deb. Also, I would like to thank Valerie L. Shalin for her expertise in appreciating always the global implications of the work. Her insight has prompted me to keep in mind the *context* in which individuals learn, develop, and perform and to strive for the integration of knowledge within the appropriate context. Thank you, Valerie.

Also, I wish to thank the other members of my committee: Corey E. Miller, whose continuous support helped me to persist throughout this project; David M. LaHuis, whose wry wit caused me to rethink constantly my understanding of the fit of this, or any research project; and Martin P. Gooden, whose insight led me to investigate the proposed theoretical underpinnings of stereotype threat relative to testing. Thank you all.

Additionally, I wish to thank the undergraduate research assistance who helped with this research: Teresa Jones, Jenny Jenks, Lisa Brown, Dan Divine, and Jennifer Vonderahe…and, of course, a special ‘thank you’ goes to Keith Leas, a newer Graduate student who helped me keep it all under control. Thank you all for your dedication to this project.

Finally, I would like to thank Mrs. Rose M. Schultz, Leanna M. Mellott, and especially, Jordan L. Palumbo for giving me the desire to persist during all of those times when the final goal seemed unreachable.

I. INTRODUCTION

The tests that employers use for personnel selection have an inherent risk of adverse impact (A.I.), that is, the systematic bias against a particular group within the labor market. Adverse impact in selection can be extremely costly to U. S. organizations. A determination of adverse impact in a selection system often requires the company to re-examine existing methods and develop and/or implement alternative devices. Further, a finding of adverse impact can force the company to pay out millions in damages. For example, in the case of Robinson v. Ford Motor Company, Inc. (2004), in response to a finding that their Apprentice Training Selection System test (ATSS) unfairly discriminated against African Americans, Ford offered a settlement that included suspension of their existing selection system, the hiring of an independent expert to "design and validate an apprentice selection instrument, or instruments" (p. 5), and "pay all reasonable costs and expenses associated with the work of the Expert" (p.6). Further, Ford agreed to pay $2,400 to each of the 3,420 'Settlement Class Members' (= $8,208,000), each of the thirteen 'Class Representatives' and 'Charging Parties' the sum of $30,000 (= $390,000), and $1,100,000 to the 'Settlement Class Counsel' with an additional $567,000 to be awarded over the next three to five years (= $1,667,000). The final cost to Ford was nearly $10 Million, not including the costs associated with the company's own legal fees, employee time and effort, and possible productivity lost due to managing this lawsuit.

As can be seen from the above example, lawsuits associated with adverse impact are expensive. Moreover, the Ford lawsuit is but one that has been filed with the courts in response to poor selection methods. The predictor used by Ford, its Apprentice Training Selection System (ATSS) test, was found to be discriminatory, resulting in disparate treatment against African Americans. Although using the ATSS may have been effective in predicting training performance, the use of alternative selection methods, or the identification of effective, alternative predictors may have avoided the lawsuit or affected the court's decision. In short, the use of alternative predictors may help protect companies from the high costs that result from discrimination lawsuits. Thus it is important to examine and extend existing research on prediction in order to identify effective predictors that reduce incidence of unfair discrimination, i.e., adverse impact.

Research has demonstrated that cognitive ability is an effective predictor of performance. However, the mechanisms through which cognitive ability affects performance have yet to be identified completely, and these mechanisms might prove to be better predictors of performance and result in less adverse impact. Schmidt and Hunter (1998) speculated that cognitive ability affects performance indirectly through the acquisition of job knowledge. Therefore, one purpose of this study is to test job knowledge as a mediator of cognitive ability effects on performance. However, the strength of this mediation effect may depend on how we define and measure job knowledge (Palumbo, 2004). More specifically, current definitions of job knowledge may limit our ability to assess adequately the influence of job knowledge on performance. Therefore, a second purpose of this study is to extend current definitions of job knowledge and to test different measures of job knowledge as mediators of the cognitive ability - performance relationship. Furthermore, it may be the case that job knowledge is actually a better predictor of performance than cognitive ability (Palumbo, Miller, Steele-Johnson, & Shalin, 2005). Again, this may depend on how job knowledge is defined and measured (Palumbo, 2004). Thus, a third purpose of this study is to compare the efficacy of cognitive ability and these different types of job knowledge tests in predicting performance.

Despite its role as an effective predictor of performance, cognitive ability tests often results in subgroup score differences, thereby resulting in adverse impact on selection (Farr, 2003; Hough, Oswald, & Ployhart, 2001; Sackett & Ellingson, 1997). Research is underway to identify possible causes of the subgroup test score differences (e.g., Ellis & Ryan, 2003; Hausdorf, LeBlanc & Chawla, 2003). One such factor that has been proposed is stereotype threat (Steele & Aronson, 1995; Steele, 1997). If job knowledge is equal to cognitive ability in predicting performance, and if job knowledge tests demonstrate smaller subgroup score differences relative to cognitive ability tests because of the absence of stereotype threat that affects cognitive ability tests, then job knowledge tests might decrease the instances of adverse impact and offer a viable alternative to the use of cognitive ability tests. Therefore, the final purpose of this study is to provide a preliminary examination of stereotype threat as it relates to cognitive ability and job knowledge tests.

Cognitive Ability

Cognitive ability test purpose. Cognitive ability tests are used often in employee selection procedures due to their low cost and ease of administration. More importantly, apart from administrative efficiency, cognitive ability tests have been seen as effective predictors of performance, having a validity coefficient ranging from .23 for unskilled (labor) jobs to .58 for 'professional-managerial jobs' (Hunter & Schmidt, 1996; Schmidt & Hunter, 1998). Therefore, cognitive ability tests are widely used in industry for initial job candidate selection and selection for entry into job-related training programs.

The nature of the construct of cognitive ability. "Cognitive ability" is the term used to describe a general ability factor. This factor is operationally defined as the common variance shared by tests of specific abilities (Hough & Oswald, 2000). More specifically, tests of various specific (mental) abilities (e.g., spatial ability, verbal ability, numerical ability, etc.) are positively intercorrelated, and the mental ability that is reflected by that shared variance is labeled *cognitive ability*. Sometimes referred to as *g*[1], (e.g., Ree & Earles, 1991; Ree, Earles, & Teachout, 1994), or Spearman's *g* (e.g., Conway, Kane, & Engle, 2003), IQ (e.g., Rushton & Jensen, 2005), or general mental ability (GMA) (e.g., Schmidt & Hunter, 1998; Schmidt, Hunter, & Outerbridge, 1986), this common variance is identified as "the general factor common to diverse cognitive tests" (Jensen, 1995. p.42). Furthermore, this general factor accounts for a greater proportion of the overall variance in all of the tests than any other factor (Nyborg & Jensen, 2000).

Cognitive ability as intelligence. Cognitive ability is often viewed as intelligence. This view is evident in statements made by researchers throughout existing literature. For example, in her discussion regarding 'cultural equivalence' of cognitive ability tests (CATs), Helms (1992) stated that in her view, "CATs refers to those measures designed to assess intelligence, mental abilities, cognitive abilities, and scholastic aptitude because,..., these terms are virtually synonymous." (p. 1083). Also, in his meta-analysis of the literature regarding performance prediction, Hunter (1986) stated that "'general cognitive ability' is used here in place of the term 'intelligence'" (p. 341). Thus, a common theme that extends through the literature is that general cognitive ability and general intelligence are different labels for the same construct.

Intelligence is described as an individual difference factor relating to one's ability to understand complex ideas, adapt to one's surroundings, learn from experience, engage in reasoning, and surmount obstacles mentally (Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg, & Urbina, 1996). Additionally, intelligence is considered to be "essentially stable in its nature throughout life" (Vernon, 1979, p. 8). Neisser et al., (1996) agree with this perspective of the stability of intelligence, as evidenced by their ongoing discussion of intelligence as a stable, dispositional trait. Thus, if cognitive ability is intelligence, and intelligence is a relatively stable, dispositional trait, then cognitive ability also must be viewed as a relatively stable construct.

Relationship to performance. As stated earlier, cognitive ability tests are commonly used in selection because of their relative effectiveness for predicting performance. Regarding attempts to improve validity in predicting job performance, Tenopyr (2002) wrote that, "Emphasis on g appears to have been the most fruitful approach." (p. 113). Thus, researchers have focused on demonstrating the effectiveness of cognitive ability (g) as a predictor (e.g., Hunter, 1986; Schmidt & Hunter, 1998). The results of much of this research have demonstrated that cognitive ability is the best, *general* predictor of performance across a variety of jobs (Gottfredson, 2002; Hunter & Schmidt, 1996; Schmidt & Hunter, 1998). In fact, Hunter (1986) went so far as to say that "general cognitive ability predicts job performance in all jobs" (p. 340) although he noted also that job complexity affected the predictive validity of cognitive ability tests. More specifically, in his meta-analysis of hundreds of studies examining cognitive ability effects on performance ratings, Hunter (1986) reported a validity coefficient of .58 for high complexity jobs, .51 for medium complexity jobs, and .40 for low complexity jobs.

Further, for individuals without prior experience on the job, general mental ability (GMA) is the most valid predictor of future performance *and learning* (Schmidt & Hunter, 1998). Specifically, general mental ability has been shown to predict "performance (amount learned) in job training programs" with an average predictive validity of .56 (Schmidt & Hunter, 1998, p. 266). This statement is important because it addresses not only the role of cognitive ability in selection for future performance, but more importantly, it addresses also the role of cognitive ability in learning. Specifically, not only does cognitive ability predict job performance, but it effectively predicts also the acquisition of job-relevant knowledge. Therefore, cognitive ability tests are seen as effective for choosing among individuals with no previous job knowledge to be assigned to formal training programs.

Relationship to adverse impact. One limitation of the use of cognitive ability tests in selection is that these tests often result in subgroup score differences, thereby resulting in adverse impact on selection of entire subgroups (Farr, 2003; Hough, Oswald, & Ployhart, 2001; Sackett & Ellingson, 1997). More specifically, observed scores for blacks are, in general, approximately one standard deviation (1 SD) lower than for whites on cognitive ability tests (Hunter & Hunter, 1984; Sackett & Ellingson, 1997). The use of cognitive ability test scores for selection decisions results in relatively fewer minorities hired, i.e., adverse impact (Hough, Oswald, & Ployhart, 2001). Thus, the use of cognitive ability tests for selection often results in adverse impact.

Title VII of the Civil Rights Act (1964) and the Tower Amendment (Section 703h, Civil Rights Act, 1964) prohibit intentional discrimination in personnel decisions based on race, color, religion, sex, or national origin. Subsequent cases (e.g., Griggs v. Duke Power Co., U. S, Supreme Court, 1971) exposed the possibility of the unintentional discrimination against members of these protected groups when tests scores differ for their members from those achieved by majority members. Thus, the courts eventually defined adverse impact as ***unintentional*** discrimination centered on the negative effects of a measure or selection test for women and minority groups (U. S. Supreme Court, 1971). If a court determines that a selection system does, in fact, exhibit adverse impact, the employer must demonstrate that the selection system is job relevant and that business necessity exists for its use (Aramburu-Zabala Higuera, 2001).

To address the issue of possible adverse impact in selection testing, researchers have begun to examine alternatives to the traditional use of cognitive ability alone as a predictor of performance, such as the use of multiple predictors in the selection process or the inclusion of alternative measures of performance in the criterion. Furthermore, despite Sackett and Ellingson's (1997) demonstration of the limited benefit of using multiple predictors, other researchers have continued to call for the inclusion of even more predictors (both cognitive and non-cognitive) in the selection battery and/or the further expansion of the criterion domain (e.g., Goldstein, Zedeck, & Goldstein, 2002; Helms, 2002).

In terms of finding better sets of predictors, researchers have begun to attempt to identify factors that may have an effect in the race – cognitive ability testing relationship. For example, Ellis and Ryan (2003) examined the mediating effects of test preparation, test-taking strategy use, and self-efficacy in the relationship. They found that although blacks reported having received more test preparation training than whites, reported using more test-taking strategies than whites, and reported high levels of test-taking self-efficacy, blacks scored lower than whites on a cognitive ability test. Ellis and Ryan attributed the lower scores to the greater use of *ineffective* strategies by blacks than by whites.

Researchers have examined also the efficacy of using alternate types of cognitive ability tests. For example, to examine the question that the mean score differences may be caused by a reading comprehension factor inherent in certain cognitive ability tests (e.g., the *Wonderlic*), Hausdorf, LeBlanc, and Chawla (2003) compared performance on the General Aptitude Test Battery (GATB) to that on the Raven's SPM. Hausdorf et al. found that both tests resulted in similar levels of black-white score differences, and therefore, the use of either type of cognitive ability test would result in adverse impact against minority candidates.

Finally, researchers have begun to attempt to identify alternative approaches to assessment that would reduce or eliminate the occurrence of adverse impact in selection. One approach that has been explored in the selection literature is the use of multiple predictors for hiring decisions. For example, Helms (2002) suggested the use of "cultural identification" measures along with tests of cognitive ability. Also, Goldstein, Zedeck, and Goldstein (2002) supported the idea of including even more cognitive and non-cognitive predictors in the selection battery. To this end, researchers have examined the effectiveness of adding other predictors to tests of cognitive ability, thereby creating a composite for the purpose of reducing subgroup differences (e.g., Baron & Janmen, 1996; Sackett & Ellingson, 1997). One approach has combined predictors in assessment centers (ACs). Typically, assessment centers are designed as simulations of the work environment where trained raters observe and evaluate applicants' behavior in an attempt to successfully predict future performance. (See Cascio, 1998, for a detailed discussion of assessment center methodology). Baron and Janmen (1996) found that in assessment centers, test scores differences exist between racial groups but that the differences were not as large as those found with cognitive ability tests. However, in their meta-analysis of these efforts, Sackett and Ellingson (1997) demonstrated the limited effectiveness of adding more predictors alone. More specifically, these authors found that most of the change in the relationship between predictor and subgroup score difference occurs when the first two or three predictors are added to cognitive ability tests to create the composite. The addition of further (4 or more) predictors provides diminishing gains, if any.

Another approach to reducing adverse impact seemed to redefine testing criteria or use other types of tests. For example, Barret, Carobine, and Doverspike (1999) were able to demonstrate reduced black-white subgroup differences by using short term memory (STM) tests. In this case, whereas cognitive ability tests produced a subgroup score difference of .80, the difference on the short term memory test was .39. Furthermore, Verive and McDaniel (1996) demonstrated that the scores on short term memory tests and job performance were highly correlated (*r* = .45), thereby providing evidence for the validity of these measures.

Additionally, there is evidence that work sample tests reduce adverse impact (Aramburu-Zabala Higuera, 2001). Moreover, work sample tests have been found to have predictive validity that is equal to or greater than other, paper and pencil measures (Aramburu-Zabala Higuera, 2001; Cascio & Phillips, 1979). Similarly, Schmidt & Hunter (1998) demonstrated that work sample measures were slightly more valid for predicting performance than measures of general mental ability [correlations with performance (*r*s) = .54 and .51 for work sample measures and general mental ability (GMA) measures, respectively].

An important point here is that the use of work sample tests represents a 'shift' from the practice of using 'signs' of ability for selection to the use of 'samples' of job-related behavior to predict future job performance. More specifically, rather than using general cognitive ability (a well-established sign of general ability and one that is inexpensive to administer measures of) to predict performance, this line of research has focused on actual samples of work and the knowledge that guides decision-making about the job. There is evidence to suggest that work samples produce lower adverse impact and provide reliable predictors of both training performance and job performance (Hough & Oswald, 2000). Thus, a work sample approach to adverse impact mitigation merits further exploration.

This 'samples' approach moves testing closer to issues of content and job-related understanding and is consistent with legal requirement to relate testing procedures to job requirements. Wernimont and Campbell (1968) based selection on a work sample model, proposing the idea of behavioral consistency, i.e., that prior performance predicts future performance. These authors argued that in attempts to predict job behavior “it will be much more fruitful to focus on meaningful samples of behavior, rather than signs of predispositions, as predictors of later performance.” (p. 372). Thus, researchers have implemented the use of situational tests as a measure of existing judgment that could be used to predict decision-making in the future (e.g., Lievens & Coetsier, 2002). Situational judgment tests require the individual to make decisions based upon their understanding of the information provided, the declarative content of the task, and their pre-existing knowledge regarding this type of task or scenario. More specifically, situational judgment tests place individuals into an hypothetical problem scenario and require them to resolve the problem. Resolution requires individuals to *apply* their existing knowledge above and beyond the task content provided. In implementing this idea as it related to adverse impact, Lievens and Coetsier (2002) were able to demonstrate that situational tests added 3.1% predictive validity over cognitive ability tests.

In the current study, I focused on the use of sample measures (job knowledge) rather than sign measures (cognitive ability) as predictors of performance, in order to examine black-white differences. Thus, this study was designed to extend existing research using this ‘behavioral consistency’ approach by examining the effectiveness of using ‘sample’ measures other than situational tests. Further, this study will extend previous, preliminary research (e.g., Hunter, 1993; Schmidt, Hunter, & Outerbridge, 1986; Palumbo, 2004; Palumbo, Miller, Steele-Johnson, & Shalin, 2005) that has demonstrated that job knowledge mediates the relationship between cognitive ability and performance.

In summary, despite much research demonstrating the possibility of alleviating adverse impact, cognitive ability tests continue to be used in selection and the use of cognitive ability tests continues to result in significant black-white mean score differences. Regarding black-white mean test score differences, Farr (2003) wrote, "These mean test performance differences have been quite resistant to change, despite the number of approaches that have been attempted" (p. 179). Furthermore, according to Sackett and Ellingson (1997), these "commonly observed black-white score differences produce violations of the four-fifths rule at all but the very highest selection ratios" (p. 712). Thus, continued use of cognitive ability tests in predicting performance is likely to be paralleled by continued adverse impact. Therefore, it seems important to continue the search for alternative tests that reduce adverse impact and to identify factors that may contribute to these score differences, and alleviate the cause of these differences, if possible.

Job Knowledge

Job knowledge test purpose. In industry, paper-and-pencil job knowledge tests are used for candidate selection, job placement, training assignment decisions, and organizational advancement considerations. Paper-and-pencil job knowledge tests are seen as practical for measuring job performance because of the relatively low cost and ease of their administration. Additionally, these paper-and-pencil tests are correlated with hands-on measures of job performance, although the correlations are not strong enough to support the use of paper-and-pencil job knowledge tests as a substitute for hands-on measures, i.e., *physical* job samples (Dubois & Shalin, 1995). Further, similar to hands-on measures, paper-and-pencil job knowledge tests are considered to be samples of the task(s) performed on the job. Thus, written tests are designed to reflect the requirements for successful job performance. The degree of content validity is a measure of the amount of overlap between test content and the job.

However, the use of paper-and-pencil job knowledge tests rather than the use of cognitive ability tests may increase ultimately the cost of the selection system. More specifically, although inexpensive to administer, paper-and-pencil job knowledge tests may be expensive to develop, depending on the nature of the information to be assessed. Thus, the use of job knowledge tests may increase organizational costs beyond that realized by the use of existing, marketed cognitive ability tests. Therefore, the use of job knowledge tests is dictated only to the extent that they offer better prediction.

The nature of the construct of job knowledge. Job knowledge may be viewed as knowledge within the context of, and relevant to, the job. Simply put, we are addressing the construct of knowledge in general but as it applies to the specific context of work. This general construct (knowledge) has been researched thoroughly over the past few decades by cognitive scientists (e.g., Greeno, 1989). Thus, to understand job knowledge, we must begin with an examination of the literature on knowledge that exists in cognitive psychology.

Cognitive literature on knowledge. In the literature from cognitive psychology, job knowledge is viewed as a multi-faceted construct. Cognitive researchers have identified many specific types of knowledge. These separate knowledge types interact with one another to contribute to task success. Thus, job knowledge, as a whole, may be viewed as a combination of these distinct yet interacting knowledge components. Therefore, all of these distinct types of knowledge should be valued by industrial/organizational (I/O) psychologists interested in the development of job knowledge tests.

Initially, cognitive researchers focused on identifying the basic types of knowledge that were associated directly with the content of the task. Early research identified two types of task-related knowledge: declarative and procedural. Anderson, Greeno, Kline, and Neves (1981) described declarative knowledge as knowledge of the factual, descriptive information associated with the task. Additionally, Anderson et al. (1981) described procedural knowledge as the knowledge involved in executing the task processes. More specifically, according to Anderson (1997) (who, incidentally, seems to be one of the few cognitive researchers who continues to focus on declarative and procedural knowledge to date), procedural knowledge specifies how we use declarative knowledge to solve problems. Thus, procedural knowledge is knowledge of how to do the task and is therefore essential in task completion. Further, in their discussion of knowledge acquisition, Anderson et al. (1981) stated that these two types of knowledge are acquired relatively early in the (task) learning process (to be discussed later in more detail as it relates to the industrial/organizational conceptualization of job knowledge).

However, cognitive researchers, including some of those mentioned above, seemed to recognize that this early view of job knowledge may have been limited because it addressed only some of the information (knowledge) needed to successfully perform a task (e.g., Greeno, Riley, & Gelman, 1984). More specifically, this view of job knowledge included only the identification of knowledge of a set of actions or subtasks to be accomplished within a given task and the procedures used to accomplish those actions. Thus, this view concentrated only on the ‘Whats?’ and ‘Hows?’ that are included in the task. Many modern jobs have evolved into dynamic activities that often involve shifting our focus, prioritizing our actions, and making appropriate decisions. In this arena, *adaptability* is often the key to success. Therefore, researchers began to expand their efforts to include also the identification of the more qualitative components of job knowledge that enable adaptation.

A more general view of the combination of declarative and procedural knowledge is ontological knowledge. Ontology refers to the objects in a task and their relationships with one another (Greeno, 1989). Thus, ontological knowledge is the understanding of the relations between task elements and how each element affects the others.

Additional qualitative task knowledge components that have been identified in the literature include explanatory knowledge, tacit knowledge, and goal-recognition knowledge. Explanatory knowledge supplies the justification for a procedure to be used within a given task (Greeno, Riley, & Gelman, 1984). Tacit, decision-making knowledge is knowledge that is "not openly expressed or readily stated" (Hedlund, Forsythe, Horvath, Williams, Snook, & Sternberg, 2003, p. 117; Schmidt & Hunter, 1993) and is believed to be acquired directly from the task and to increase with experience (Sternberg, 1985). Finally, goal-recognition knowledge associates the recognition of task outcome requirements with the understanding of the processes required for successful completion (Greeno & Simon, 1988).

All of these qualitative knowledge components contribute to task performance by supplying task-relevant information above and beyond the 'Whats?' and 'Hows?' of the task. Therefore the identification of these components allows us to begin to identify also the individual's knowledge of why and when to do which task procedure. Deciding how to prioritize and when and how to implement subtasks often involves planning, an awareness of existing resources and their ‘appropriate’ allocation, an awareness and understanding of existing conditions, an understanding of the interrelatedness of all parts of the task, and a focus on future goals. Thus, the ‘Whens?’ of a task seem to be driven by these different yet integrated qualitative types of knowledge.

Finally, and perhaps most importantly, it seems that our conceptualization of job knowledge must include knowledge of the “Whys?” of a task. It is the 'Whys?' that determine our choice of which action or procedure to implement. Knowing the ‘Why?’ allows us to prioritize subtasks, to determine choices. Thus, knowing the ‘Why?’ often allows us to determine the ‘When?”. Knowing the ‘Why?’ allows us to reason about a task and/or reason through serial task requirements. Knowing the ‘Why?’ allows us to solve problems. Knowing the ‘Why?’ allows us to search for analogous instances or situations. Therefore, knowing the ‘Why?’ allows us to generalize our knowledge, to recognize similar situations and implement previously successful strategies. In short, knowing the ‘Why?’ allows us to succeed in the job. Therefore, knowing the ‘Why?’ *must* be included in any conceptualization of the knowledge that is necessary to do the job, or ‘job knowledge’. The qualitative components of job knowledge that have been identified in the cognitive research afford that necessary inclusion because they do address the ‘Whys?’ that drive job performance.

Furthermore, researchers have begun to realize that the acquisition of job knowledge occurs within a given, specific situation and more recent literature has included this idea (e.g., Greeno, 1989; Lave & Wenger, 1991).

Job knowledge as malleable (vs. stable). In the cognitive literature, knowledge is viewed as malleable, an outcome of the ongoing process of learning. Anderson et al. (1981) described the learning process as iterative wherein individuals acquire declarative knowledge first and then procedural. Also, research has demonstrated that throughout the learning process, individuals were able to acquire increasing levels of various types of competencies (due to increasing knowledge) regarding a particular task (e.g., Greeno, Riley, & Gelman, 1984; Smith, Greeno, & Vitolo, 1989). Similarly, task-related tacit knowledge is believed to increase with experience (Sternberg, 1985). Furthermore, research on expert-novice differences has demonstrated that, as more learning occurs, individuals categorize problems more effectively and reason more effectively about their solutions (e.g., Chi, Feltovich, & Glaser, 1996; Van Lehn, 1996). Thus, there is ample evidence that knowledge changes over time and is therefore malleable, and presumably, this malleability extends to all types of knowledge.

Current industrial/organizational conceptualizations of job knowledge. In industrial/organizational research and practice, our conceptualization of job knowledge comes from our conceptualization of work. Our conceptualization of job knowledge, then, drives how we operationally define job knowledge. We have derived much of our current concept from research in cognitive psychology. However, our current concept may be limited because we may have overlooked some of the relevant constructs that cognitive research has to offer. As industrial/organizational psychologists, we must consider also cognitive research that addresses types of knowledge that contribute to our ability to adapt our thinking. In short, whereas job knowledge is conceptualized quite broadly in the cognitive literature, the industrial/organizational concept is much narrower.

Therefore, we should continue to borrow from our cognitive colleagues in order to extend our conceptualization of job knowledge and thus, our definition.

Existing industrial/organizational operational definitions identify job knowledge as technical information, facts, and procedures required to do the job (e.g., Hunter, 1993; Schmidt, Hunter, & Outerbridge, 1986). Additionally, industrial/organizational researchers have used variations of the concept of job knowledge. For example, Schmidt et al. (1986) assessed job knowledge through "written measures of facts, principles, and so forth, needed to perform the job" (p.433). Additionally, in his examination of the effects of conceptions of ability on learning, Martocchio (1994) limited his assessment of the knowledge that was acquired during training to that of declarative knowledge only. Also, in their examination of goal orientation and ability effects on knowledge, Bell and Kozlowski (2002a) limited their basic knowledge test to one that assessed declarative knowledge. Similarly, Fisher and Ford (1998) investigated the effects of individual effort and goal orientation on knowledge. However, as they described, their knowledge measure focused on “facts found in the text of” the learning materials provided (p. 407). Thus, this particular knowledge test assessed also declarative knowledge only. Thus, in practice, industrial/organizational psychologists have focused, at least partially, on declarative knowledge ("facts") in their operational definition of job knowledge. Furthermore, the approach of focusing on facts supplied in training materials used by Fisher and Ford (1998) also puts the burden of task analysis on trainers. Thus, Fisher and Ford’s (1998) approach may be risky, because, in this case, these trainers may be likely to identify only those aspects of knowledge that lend themselves to formal training.

Hunter (1993) extended the concept of job knowledge beyond the declarative by defining job knowledge as "knowledge of the technical information about objects and concepts required to do the job and knowledge of processes and judgmental criteria required for efficient or correct action on the job" (p.258). Thus, according to Hunter's (1993) definition, job knowledge consists of a combination of declarative *and* procedural knowledge. Similar definitions can be found elsewhere in the literature. For example, in their discussion of the effects of adaptive guidance on several training outcomes, Bell and Kozlowski (2002b) refer to basic (task) knowledge as "the fundamental principles and operations of a task" (p. 273). Yet, their measure of basic knowledge was a composite of these two knowledge types. Thus, it seems that basic knowledge, also, is seen as a combination of declarative and procedural knowledge only.

Thus, industrial/organizational practitioners have a well-established history of focusing on these two types of knowledge. However, to reiterate, this view of job knowledge may be limited, because of its limited focus on knowledge of a particular set of actions and/or the procedures used to accomplish those actions. Again, in today's jobs, *adaptability* is a requisite for success, but a concept that we in industrial/organizational psychology have yet to consider fully. Therefore, our concept of what work is, and thus what job knowledge is, may be out-dated. We seem to view work as a series of tasks to be performed while on the job. Therefore, we continue to concentrate on the 'Whats?' and 'Hows?' that are included in the task. This is an important first step, because the 'Whats?' and 'Hows?' are imperative for task success and therefore *must* be considered an integral part of our concept of job knowledge. However, it may be the case that we often omit the 'Whens?' of a task from our thinking.

An exception to this scenario involves the type of task in which success requires subtasks to be performed in a specified sequence or order. In this type of (often static) ordered task, the 'When?' refers to the order of operations for subtasks to be carried out. For example, when performing Cardio-Pulmonary Resuscitation (C.P.R.), all subtasks must be accomplished in the correct order for the task to be successful.[2] In this type of task, the 'When?' refers to the correct "ordered set" of procedures or subtasks, and the knowledge of the procedural set would be included in our (industrial/organizational) overall concept of job knowledge. In sum, we in industrial/organization psychology are accomplished at including the 'Whats?', 'Hows?', and, in specific, ordered cases, even the 'Whens?' of a task in our conceptualization of job knowledge.

However, as discussed earlier also, this set of facts and procedures may not reflect the complete nature of the job requirements. If the task is dynamic and involves discretionary decision-making, as is the case with many of today's jobs, knowing *when* to do which procedure may depend on other information, the 'Why?'. Thus, the 'when' seems to be driven by those different, yet integrated qualitative types of knowledge identified in the cognitive literature. These knowledge types represent the 'Whys?' of the task, and therefore these should become integrated into our conceptualization of job knowledge also.

This is not to say that industrial/organizational psychologists are not interested in the 'Whys?' of a task. In fact, recently, some research has begun to address the role of the effects of 'other' types of knowledge on performance. For example, in their examination of the relationship between goal orientation, learning strategies, learning outcomes, and the transfer of knowledge, Ford, Smith Weissbein, Gully, and Salas (1998) administered a test that attempted to assess both declarative and strategic knowledge. However, according to their description, the test consisted of only nine items, “several” (of which), according to the authors, “tapped declarative knowledge” (p. 225). Thus, their test could include very few items that “were written to assess strategic knowledge of the most efficient way to access information cues...and how to prioritize targets...” (p. 225). Also, their test did not include any assessment of procedural knowledge. Thus, it eliminated any opportunity to assess one of the important ‘basic’ building blocks of knowledge that contribute to the strategic. Therefore, although their approach was a positive attempt to assess knowledge beyond the declarative, the test used by these authors may have provided an incomplete assessment of knowledge because of its limited length and scope.

Additionally, in both studies mentioned earlier, Bell and Kozlowski (2002a, 2002b) described the possible role of "strategic" knowledge (also identified as ‘tacit’ by these authors) and included a measure of this knowledge type in their research. Bell and Kozlowski (2002b) described strategic knowledge as “the extent to which a trainee has learned the underlying or deeper complexities of a task” (p. 273). However, because their description failed to address any of the other qualitative types of knowledge that may interact to influence the strategic, their measure of this type of knowledge had similar shortcomings. Further, Fisher and Ford (1998) identified *application* (knowledge) as a second outcome of learning, and implemented a measure of application in their study. However, similar to Bell and Kozlowski (2002b), Fisher and Ford (1998) failed to capture any other knowledge types. Thus, although this research adds the idea of strategic or application knowledge to our concept of job/task knowledge, its effectiveness may be limited because it omits any assessment of the other, qualitative types of knowledge. Through this omission, we eliminate also any opportunity to identify accurately and systematically all that has been learned. This could have detrimental effects on our attempts at training evaluation, especially, by limiting our ability to determine what type(s) and how much of the intended information the individual has acquired during the training.

Thus, the current industrial/organizational definition of job knowledge may limit our ability to adequately assess the knowledge that the individual must possess to successfully perform a task because the focus of our definition may be too narrow to allow us to assess other knowledge types that influence successful task performance. For example, this definition overlooks the influence of explanatory and goal knowledge (Dubois & Shalin, 1997), among others. Further, our conceptualization of job knowledge determines how we attempt to measure the construct. More specifically, our general conceptualization of job knowledge drives how we operationally define job knowledge and this definition, in turn, acts as a guiding force in our approach to developing job knowledge tests.

Thus, in industrial/organizational psychology, the construct of job knowledge has not been well-defined, the definition being too narrow in its focus, even in some of the more recent papers (e.g., Schmidt & Hunter, 1998). Therefore, researchers should examine more fully the construct of job knowledge. Specifically, I propose that we should expand the current definition of job knowledge to encompass the individual's overall understanding of the task, that is, how to apply job-specific information for successful task completion. I suggest that research that expands the definition beyond declarative and procedural knowledge to include knowledge regarding the appropriate application of job-specific information might be more informative about the candidate's level of knowledge about the task. Potential benefits might be a better understanding of the knowledge-performance relationship, improved prediction, and/or decreased incidence of adverse impact.

Distinction between basic knowledge and understanding. Although all task-related knowledge comes from the content of the task, it is important to distinguish between various types of knowledge relative to their ability to predict performance. Therefore, for the purposes of this study, basic knowledge has been identified as a combination of declarative and procedural knowledge, and the measure of basic knowledge was a composite of these two knowledge types as described by Bell and Kozlowski (2002).

Understanding necessarily subsumes declarative and procedural knowledge in that both are necessary precursors to the decision-making process. Thus, task "understanding" is defined as the appropriate combination of declarative, procedural, ontological, explanatory, tacit, and goal-recognition knowledge. To clarify, by 'the appropriate combination' I mean that combination which allows the individual to accurately, successfully complete the task.

However, I suggest that the measure of understanding need not assess declarative and/or procedural knowledge directly because of the assumption that both knowledge types are present in the individual's knowledge repertoire. This combination can only be assessed by determining what priorities are assigned, what choices are made, and what procedures are applied. In this case, tests developed by traditional industrial/organizational methods may not capture job knowledge adequately because we have yet to include understanding as part of our conceptualization of job knowledge to date. Thus, there seems to be no existing method in industrial/organizational psychology that systematically assesses the complete set of qualitative knowledge types that I call "understanding." Therefore, for the purposes of this study, the measure of understanding was developed using Dubois and Shalin's (1995, 1997) approach. The test was designed to assess the individual's understanding of the appropriate application of the more qualitative types of knowledge relative to the task at hand.

Relationship to performance. As stated earlier, substantial research has demonstrated that cognitive ability is an effective predictor of performance (e.g., Gottfredson, 2002; Hunter & Schmidt, 1996; Schmidt & Hunter, 1998). There has been less research examining job knowledge effects on performance. However, these studies have demonstrated that job knowledge is an effective predictor of performance (e.g., Hunter & Schmidt, 1996; Palumbo Miller, Steele-Johnson, & Shalin, 2005; Schmidt & Hunter, 1998). Moreover, much of the existing work has examined job knowledge in the context of mediation. For example, Schmidt and Hunter (1998) speculated that cognitive ability affects performance indirectly through the acquisition of job knowledge. Other existing research has demonstrated that job knowledge functions as a mediator between cognitive ability and its effect on performance (Hunter, 1993; Hunter & Schmidt, 1996, Schmidt & Hunter, 1998). However, in these studies, job knowledge only partially mediated the effects. Furthermore, Schmidt et al (1986) suggested that the job knowledge effect was moderated by the sample, observing stronger effects in military samples than in civilian samples, perhaps due to the military's emphasis on procedures.

More recently, job knowledge has been shown to completely mediate cognitive ability effects on performance and account for significantly more of the variance in performance than cognitive ability (Palumbo et al, 2005). The apparent difference in these findings was related to how job knowledge was defined, and ultimately measured (Palumbo, 2004). Thus, the ability of job knowledge to mediate cognitive ability effects on performance may depend on how we measure job knowledge.

Relationship to adverse impact. As stated earlier also, job relevance of selection tests is one requirement of adverse impact legislation. More specifically, if a selection test/battery is determined to be discriminatory, the organization must provide evidence that the information addressed by the test is related to the job for which it is used. Job knowledge tests are clearly job related because they are samples of the knowledge that is required for job success and thus are designed to effectively predict job performance. Therefore, the use of job knowledge tests ('samples') represents a more legally-defensible selection device than the use of cognitive ability tests ('signs') to predict performance.

Test Design

The assessment of job knowledge requires the creation of job-specific measure(s)/test(s). Different approaches to test development can be found in the literature. Generally speaking, current industrial/organizational job knowledge test development procedures may be limited by current conceptual and operational definitions of job knowledge. Therefore, I suggest that it may be beneficial for industrial/organizational researchers to examine an alternative to the current approach to job knowledge test development that may extend our definition of job knowledge.

Specifically, I propose that a cognitively-oriented approach to test development will provide a more complete picture of job knowledge.

Existing industrial/organizational test-design methods. Traditionally, in industrial/organizational psychology, the process of constructing job knowledge tests begins with an analysis of the job. In this analysis process, subject matter experts (SMEs), usually job incumbents and/or job supervisors, identify individual tasks, behaviors, and abilities that are necessary for successful task performance. (See Cascio, 1998, for a detailed discussion of the test development process.) This identification process results in the definition of the task domain by identifying categories of independent dimensions of successful job performance. Additionally, the analysis process defines the categories of types of knowledge required for task performance, i.e., the job knowledge domain. Next, the identified task elements are rated as to their importance and frequency of use in the job. This process of rating possible test items may be completed by the same group who generated the original list or by another group of job incumbents/supervisors. Those items that meet pre-established criteria are included in the written test.

Job incumbents help to develop and design job knowledge tests. These incumbents, however, have already achieved some level of expertise within the task domain and the job knowledge domain. Incumbents identify the job knowledge required for task completion based upon their own levels of experience and understanding.

Possible benefits of the existing content-oriented approach to test design. The obvious benefit to existing industrial/organizational approaches to test design is that they provide a method for assessing declarative and procedural knowledge. Both of these knowledge types have been related to task performance. Thus, the existing approach offers a good beginning in our attempts to predict that performance.

Possible limitations of existing content-oriented approaches to test design. Again, these test development methods usually involve the identification and rating of appropriate test items by subject matter experts who have ample experience on the job. However, learning is viewed as an iterative process of knowledge acquisition, understanding, and refinement (Anderson et al, 1981). Increased experience would afford an individual more opportunities to acquire more task-relevant information. This, in turn, would add to existing task knowledge and allow for increased understanding of the task and refinement of the knowledge about successful task completion. Therefore, standard industrial/organizational test design methods may not adequately assess the amount and types of knowledge that is acquired at different stages of the learning process.

Further, Landy and Vasey (1991) found that task frequency ratings were significantly influenced by incumbent's level of experience. Thus, the content-oriented tests developed by these individuals may reflect the task requirements as understood by individuals with similar levels of experience. More specifically, the tests include declarative items such as rules and facts that are necessary for task completion and procedural information that identify how to complete task processes but may not include items that assess the concepts, understood by incumbents, which inform the individual about how to use that declarative and procedural task knowledge to successfully perform the task. This information is excluded from domain definitions as identified by job incumbents. Therefore, the defined task and knowledge domains may omit necessary content information and may omit methods for representing and assessing the qualitative information about knowledge organization and use (Dubois & Shalin, 1995).

Additionally, content-oriented written tests may not provide the context to facilitate the identification of the task performance features that discriminate between levels of competence. Levels of task competence may reflect individual levels of task knowledge. For example, novices must necessarily have less task knowledge than incumbents (who help design the job knowledge tests) and therefore would perform at lower levels than those with more task expertise. Although test items are rated for importance and salience for the job, these items may be limited for identifying and sampling the diagnostic job knowledge (Dubois & Shalin, 1995). That is, items rated for importance and salience are not necessarily items that distinguish among levels of expertise.

Cognitively-oriented approach. As discussed earlier, evidence from the cognitive literature demonstrates that job knowledge includes more than declarative and procedural task information. That is, individuals must know also how to *apply* those types of knowledge to successfully perform the task. Procedural knowledge fails to supply this application information because it is limited to knowledge of how to complete those actions that are required *within* the task. Thus, procedural knowledge fails to address abstract processes that determine when and why to perform those task-specific actions. Cognitive research has shown that other knowledge is required that will allow the individual to apply the givens of the task. Therefore, I am suggesting that as researchers we should expand our conceptualization of job knowledge to include other components of knowledge, the combination of which comprises task “understanding”. The development of an alternative definition of job knowledge will enable an alternative to existing types of job knowledge tests.

Cognitive test-design methods. An alternative method for approaching test design has been described that uses observations of job expertise as a guide for analyzing the task. This cognitively-oriented approach emphasizes the description of job expertise and analysis of how experts perform tasks within the context of the job. One purpose of this type of approach is to identify more abstract procedures that are needed to accomplish task goals, abstract information that is omitted in the industrial/organizational approach. Additionally, the cognitively-oriented approach is designed to identify unique types of information, including abstract cognitive process information, which is used to guide successful execution of the task. (Dubois & Shalin, 1995; Dubois, Shalin, Levi, & Borman, 1997). That is, a cognitively-oriented approach is aimed at capturing the more abstract information, understood by more expert individuals, about the use of declarative and procedural task information in successful task performance.

Dubois and Shalin's (1995) overall approach includes three stages of implementation: knowledge elicitation, knowledge representation, and test specifications. Eliciting knowledge is accomplished through the use of verbal protocols and coaching, asking leading, task-relevant questions. Knowledge is represented using a plan-goal graph, a 'picture' that organizes task-relevant information at various levels of abstraction. Test content is specified by sampling different components of knowledge that are identified as necessary for successful task completion. These components are translated into questions, which are included in the initial item pool. The final test version is created through the (piloting) process of item administration, statistical evaluation, and refinement. As in content-oriented test development, those items that meet pre-established criteria are included in the written test.

Possible benefits of a cognitively-oriented test design. A cognitively-oriented approach to designing job knowledge tests may offer a substantial amount of information that is excluded from content-oriented tests. Tests that are content-oriented include rules or facts that govern the task whereas cognitively-oriented tests include facets of knowledge about processes that enable the individual to use those facts. The identification of unique task process information would help to define task content more completely.

Additionally, the cognitively-oriented approach allows us to identify and separate systematically the more qualitative components of knowledge content, as well as the declarative and procedural. Identification of these components of knowledge may be crucial for explaining an individual's ability to understand how to perform the task. This understanding increases as more details of the task are learned through experience. Therefore, identification of these knowledge components may be beneficial in discriminating the amount of task learning that has occurred. Further, I suggest that it is the combination of these other components of knowledge, along with declarative and procedural knowledge that the industrial/organizational content-oriented approach uncovers, that can be identified as task *"understanding"*, and that this understanding is a more effective measure of job knowledge than content knowledge, and thus a better predictor of performance. All of these components of knowledge are identified in the cognitively-oriented approach to test design.

Furthermore, a cognitively-oriented approach provides a method for representing task knowledge. This representation allows the researcher to determine the aspects of knowledge that are important for inclusion in the task description. This offers the opportunity to include substantial information about task understanding that is omitted in the content-oriented approach.

Finally, a cognitively-oriented approach provides a method for incorporating these task knowledge variables to be included in the job-knowledge test. If the ultimate goal of job knowledge tests is to determine an individual's ability to perform a task, then these tests should identify the processes that are understood about how to implement the givens of a task in order to achieve successful performance. This requires the identification of knowledge, including but going beyond the declarative and procedural information included in content-oriented tests. The resulting cognitively-oriented test may be a better tool for assessing task knowledge than the content-oriented test because the cognitively-oriented test focuses on task understanding in addition to task rules, factual, and procedural information. Thus, the inclusion of 'other' knowledge items may result in a better idea of actual task performance, and therefore, may help the researcher to better understand the relationship between task knowledge and successful task performance. Therefore, in the current study, a measure of task understanding was developed using a cognitively-oriented approach.

Summary of approaches to test-design.

In summary, existing industrial/organizational methods for job knowledge test design use a content-oriented approach. The resulting job knowledge tests identify declarative and procedural knowledge that is applicable to the task at hand. However, these tests do not include items that assess an individual's level of task understanding.

Thus, this design method may be inadequate for the measure of overall knowledge because different types of knowledge are acquired at different stages in the learning process. Alternatively, cognitively-oriented test design methods are aimed at addressing the task understanding that the individual acquires during the learning process. This understanding includes declarative knowledge, procedural knowledge, and appropriate application knowledge. Thus, a cognitively-oriented approach may better address issues of evaluating learning as a mechanism through which learning cognitive ability affects performance.

Furthermore, these traditional industrial/organizational test design methods may exclude information that may offer a better understanding of the relationship between the level of existing or acquired knowledge and task performance. Therefore, it may be beneficial to explore alternative methods for the development of knowledge tests that assess different levels of learning and include additional information that is relevant for task performance. The current study will use distinct measures to assess 1) the amount of declarative and procedural knowledge that is acquired and 2) the individual's level of task understanding.

Stereotype Threat

Definition of stereotype threat. One factor that has been proposed as contributing to mean score differences between majority and minority group members is stereotype threat (Steele & Aronson, 1995). Stereotype threat has been defined as the perception of "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995, p. 797). More specifically, stereotype threat involves membership in some subgroup (racial/ethnic/gender) and being aware of the existence of negative stereotypes associated with that subgroup. The threat comes from the perceived risk of confirming those negative stereotypes by performing poorly. Steele and Aronson (1995) proposed that stereotype threat interferes with the *intellectual functioning* of students belonging to various minority subgroups, "particularly during standardized tests" (p. 797), and that repeated experience with failure due to stereotype threat causes minority group members to withdraw from academic pursuits. Thus, stereotype threat has been discussed as a possible *social-psychological* mechanism that mediates race/gender effects on academic performance.

The detrimental effects of stereotype threat. Farr (2003) described stereotype threat as "the pressure individuals my feel when they are at risk of confirming, or being seen by others as confirming" the negative stereotypes associated with one's group (p. 179). The awareness of the group's negative stereotype is believed to be threatening enough to have disruptive effects on performance (Steele & Aronson, 1995). In an effort to find evidence in support of this idea, Steele and Aronson (1995) conducted a series of studies in which they attempted to activate stereotype threat and examine its effects. Their findings demonstrated that stereotype threat affected *both* accuracy and performance. More specifically, when the test was presented as diagnostic of ability, black participants achieved lower proportions of questions correct over the number of those attempted as well as achieving lower overall scores than whites. Finally, stereotype threat has been shown to reduce the cognitive resources available to an individual (Schmader & Johns, 2003).

However, stereotype threat is not limited to a black-white issue. Research has extended the application to examine possible effects on other subgroups as well. This research has demonstrated that stereotype threat affects the individual performance of other subgroups (e.g., women, Latinos) on advanced, challenging cognitive tests. More specifically, research has demonstrated that stereotype threat impairs the performance of women (i.e., McIntyre, Paulson, & Lord, 2003; Martens, Johns, Greenberg, & Schimel, 2006; Marx & Roman, 2002; Pinel, 1999) and of Latinos' (Schmader & Johns, 2003) on advanced math tests.

Triggering stereotype threat. Additionally, Steele and Aronson (1995) examined possible triggers of stereotype threat. They proposed, and were able to demonstrate, that presenting a test as diagnostic of intellectual ability had this triggering effect on black (but not white) participants. More specifically, when blacks believed that they were taking a diagnostic test, they demonstrated more cognitive activation of black stereotypes. Moreover, this diagnostic condition caused blacks to generate more self-doubt about their ability than others and to avoid identification with stereotypically black preferences and traits (e.g., rap music, basketball, lazy, aggressive). Blacks were more reluctant to have their race linked to their performance, as evidenced by the finding that when stereotype threat was activated, and when given the option, 75% of black participants did not identify their race on a demographics questionnaire. Steele and Aronson (1995) argued that it may be simply the act of making race salient that is enough to trigger stereotype threat.

Mediators of stereotype threat effects on performance. Steele and Aronson (1995) proposed that stereotype threat would produce *increased anxiety*, that "apprehension over possibly conforming to the negative group stereotype" (p. 801) could mediate stereotype threat effects on performance. Although these authors found no significant black-white subgroup differences in their study, more recently, Osborne (2001) found significant differences in anxiety between whites and African Americans, Latinos, and Native Americans (ALANAS) and that anxiety partially mediated race effects on academic performance and gender effects on math performance, thus providing some general support for the hypothesis. Further, Schmader and Johns (2003) found that stereotype threat affects the math performance of women and Latinos through its effects on working memory capacity. More specifically, these authors found that stereotype threat reduces the working memory capacity of members of these two subgroups and that this reduction produced lower test scores when compared to men and whites, respectively.

Moderators of stereotype threat effects. However, the effects of stereotype threat may depend on the individual's level of domain identification, or the degree to which the individual cares about and identifies with the skills being tested (Steele & Aronson, 1995). More specifically, domain identification requires that the individual views the possession of these skills as being tied to his/own sense of self-regard (Steele & Aronson, 1995). This idea has received mixed support in the existing literature. For example, McFarland, Lev-Arey, and Ziegert (2003) found that overall, individuals who reported higher levels of domain identification achieved higher scores on cognitive ability tests than those with lower levels of domain identification. However, these authors found also an interaction between race and domain identification such that this effect existed for white participants but there was no relationship between domain identification and test performance for blacks.

Another mechanism that has been discussed as affecting the relationship between stereotype threat and performance is "stigma consciousness" (Pinel, 1999). Stigma consciousness (SC) has been defined as an individual difference that "concerns the extent to which individuals are chronically self-conscious of their stigmatized status". More specifically, stigma consciousness requires not only that the individual care about the skills being tested and view those skills as a pert of his/her 'self" but that he/she focus on the stigma related to his/her group's stereotype regarding that skill. Thus, stigma consciousness has been identified also as a moderator in the stereotype threat - performance relationship.

To investigate this moderation, using the math test from a practice GRE, Brown and Pinel (2003) demonstrated that individual differences in stigma consciousness moderated gender stereotype effects on math performance of women, such that the results revealed a stronger detrimental effect of stereotypes on performance for women high in stigma consciousness and a weaker relationship between gender stereotypes and performance for women low in stigma consciousness. Furthermore, in the low threat testing condition, stigma consciousness was unrelated to test performance.

Furthermore, some researchers have begun to examine methods to *alleviate* the effects of stereotype threat by identifying other variables that may act to reduce the effects of test perceptions. For example, Marx and Roman (2002) found that exposure to interaction with, or knowledge about a competent female role model acted as a buffer to poor test performance and to perceptions of math ability. Along similar lines, McIntyre, Paulson, and Lord (2003) found that reminding (threatened) women of the abilities or achievements of other women alleviated differences in math test performance relative to that of men. Most recently, Martens, Johns, Greenberg, and Schimel (2006) found that when [stereotype-threatened] women used self-affirmation before taking a math test, they achieved scores that were similar to non-threatened women as well as men. Finally, Aronson, Fried, and Good (2002) found that when they were able to induce students to view intelligence as malleable, both African American and white students obtained higher grade point averages, while reporting greater engagement in and enjoyment of the academic process than students in control groups.

In sum, stereotype threat results from the individual's perception of or reactions to the test. According to the theory of stereotype threat, it is these perceptions that seem to result in poorer performance for members of the stereotyped group(s). Therefore, much of the existing research has focused on those perceptions in order to alter stereotype threat effects.

Finally, as noted by Farr (2003), much of the existing evidence in support of stereotype effects on cognitive ability tests is based on manipulations designed to induce specific perceptions of the tests. That is, researchers have used manipulations to either induce stereotype threat or not and then have compared test score performance in these conditions. More specifically, researchers have presented the tests as either diagnostic of intellectual ability or simply as tasks included in studies designed to investigate problem-solving (Farr, 2003).

However, consistent with theory (Steele & Aronson, 1995; Steele, 1997), the stereotype threat effect(s) should occur “naturally” when any test is perceived as diagnostic of intellectual ability as in the case of standard cognitive ability tests. This idea has support in the existing research. For example, in the African American control group used by Aronson, Fried, and Good (2002), when not instructed to view intelligence as malleable, students achieved lower GPAs than those in the experimental group who did receive the instructions.

Furthermore, this effect should occur in any setting where there is some consequence attached to test performance, as in a real-world selection situation. This idea has received support in existing literature also. For example, although in an experimental setting, McFarland, Lev-Arey, and Ziegert (2003) found that threat influenced test scores when those scores were linked to a monetary reward. Thus, consistent with previous research, I expected to find that a cognitive ability test will produce significant black-white mean score differences.

Proposed Research

To examine possible causes of black-white test score differences, two separate studies were performed. Study One was designed to focus on the examination of specific assumptions inherent in the literature on stereotype threat effects. Specifically, the first study focused on the measure of test perceptions and their effects on test and task performance. However, this first study did not address directly the effects of perceived stereotype threat relative to the tests. Therefore, a second study was designed to assess perceived stereotype threat directly and to examine the effects of stereotype threat perceptions on test scores.

Study 1

According to the stereotype threat literature (discussed above), it may be the individual's perceptions of the test that creates these significant black-white mean score differences. Specifically, the assumption is that blacks perceive specific tests as threatening and therefore perform poorly on these tests compared to whites. If this assumption is accurate, then for blacks, perceiving a test as assessing one's intelligence may be at the root of black-white test score differences. Further, it seems plausible that if even educated psychologists view cognitive ability as intelligence, and intelligence as a stable characteristic, then laymen may have the same view also. Thus, a test that measures cognitive ability should be perceived as measuring intelligence and should, therefore, cause black-white differences in scores.

However, job knowledge tests *should not* be perceived as evaluative of intelligence or ability because knowledge (as discussed earlier also) should be perceived as malleable rather than stable, and thus should evoke no such threat. This idea is supported in the existing research involving the induction of stereotype threat wherein instructions distinguished between tests presented as diagnostic of intelligence and those presented as simple tools (tasks) used for other purposes. Moreover, I expected that individuals would be able to distinguish between tests of cognitive ability and tests of task-specific knowledge. Whereas the cognitive ability tests should be perceived as diagnostic of intelligence, the task knowledge tests should not. Rather, job knowledge tests should be perceived simply as tests of job knowledge, a malleable construct.

Therefore, black-white mean differences on the knowledge tests should be smaller than those produced on cognitive ability tests

H_{1a}: Blacks and whites will have different mean scores on a cognitive ability test.

H_{1b}: A basic job knowledge test will produce significantly smaller black-white mean score differences than a cognitive ability test.

H_{1c}: An understanding knowledge test will produce significantly smaller black-white mean score differences than a cognitive ability test.

Furthermore, I expected to replicate previous findings that knowledge predicts performance better than cognitive ability (Palumbo, 2004). Thus, I expected that both knowledge tests would predict performance better than cognitive ability. Moreover, as discussed earlier, the different techniques used to design each knowledge test produce measures addressing different information. More specifically, I suggested that understanding test would be a *better* measure of knowledge because of the quality of the information captured within the test, relative to the basic knowledge test. Therefore, understanding will be a better predictor of performance than either cognitive ability or basic knowledge.

$H_{2a:}$ Basic knowledge will account for more of the variance in performance than cognitive ability.

$H_{2b:}$ Understanding will account for more of the variance in performance than cognitive ability.

$H_{2c:}$ Understanding will account for more of the variance in performance than basic knowledge.

Finally, I expected to replicate previous findings that job knowledge mediates the effects of cognitive ability on performance (Schmidt & Hunter, 1998). However, if, as suggested by Palumbo (2004), the strength of this mediation depends on how knowledge is measured, then the specific test used to examine mediation will affect the results. That is, a better test of knowledge will produce stronger mediating effects. Therefore, I expect that, whereas basic knowledge will partially mediate the relationship between cognitive ability and performance, understanding will completely mediate this relationship (Palumbo et al, 2005).

$H_{3a:}$ Basic knowledge will partially mediate the effects of cognitive ability on performance.

$H_{3b:}$ Understanding will completely mediate the effects of cognitive ability on performance.

Test Bias and Test Fairness.

The finding that one test accounts for more variance in performance than another is not sufficient to determine that it is a "better" predictor. One must consider also the possibility of test bias. According to Guion (2000), "*Bias* refers to systematic group differences in item responses, test scores, or other assessments for reasons unrelated to the trait being assessed…" (p. 433). Further, Cascio (1998) refers to the possibility of bias in selection as a "central issue in any discussion of EEO." (p. 121). Therefore, I have included a brief discussion of some of the existing literature regarding test bias and test fairness.

First, we should distinguish between test bias, discrimination, and test fairness. Bias is a technical issue, one that Guion (2000) describes as "a distortion in statistics or measurement" (p. 433). Bias can be either positive or negative and is an artifact of the selection test itself. Discrimination is simply the act of making distinctions. It is the purpose behind the administration of selection measures. A valid selection measure is one that can discriminate accurately between high and low performers. Thus, a valid selection test is a test that "accurately discriminates between those with high and those with low probabilities of success on the job" (Cascio, 1998, p. 122). The issue is whether or not the test discriminates "unfairly." Test fairness and/or *unfairness* relates to (legally) unfair discrimination, i.e., the unequal probability of subgroup members being hired regardless of equal probability of success on the job.

The Cleary (1968) model of test fairness, also known as the regression model (Guion, 2000), discusses bias in terms of the intercept, individual regression lines, and average regression line for the two subgroups. According to this model, at test is biased if the common regression line for the two groups consistently over-predicts or under-predicts the criterion scores for members of the two subgroups. More specifically, according to this model, if the intercept for the minority group is below that of the majority group, then the common regression line would always over-predict the performance of the minority group. In this case the test would be considered "positively" biased in favor of the minority subgroup. However, if the intercept for the minority group is above that of the majority group, then the average regression line (and ultimately the test) would consistently under-predict the performance of the minority group. In this case, the test would be considered as negatively biased against the minority subgroup. Thus, according to the Cleary (1968) model, a test is biased if the common regression line predicts nonzero differences in performance for the two groups. This means that a test is fair only if the regression lines are the same for each subgroup.

The Thorndike (1971) model of test fairness argues that a test is fair if the percentage of minorities selected is equal to the percentage that would be hired based on criterion scores. This requires that the score differences on the test equal the score differences in performance. However, unless there are no subgroup differences in the criterion, this model requires the regression line of the majority group to lie above the regression line for the minority. Given that subgroup differences in performance exist, the Thorndike (1971) model requires that the test over-predict performance for members of the minority subgroup (Hunter, Schmidt, & Hunter, 1977). Thus, in terms of test bias, as discussed above, the Thorndike (1971) model *requires* the test to be biased in favor of the minority.

However, as discussed in the literature, test bias does not necessarily imply test unfairness. Whereas test bias is a measurement issue, test unfairness (or fairness) is discussed as a societal, political, and legal one (Cascio, 1998; Guion, 2000). Both have implications for unfair selection decisions (i.e., adverse impact) and therefore must be included in this discussion.

The topic of fairness of tests has been an ongoing discussion in the literature, including the distinction between bias and fairness. Specifically, the one approach to determining test fairness that is most widely accepted is consistent with the Cleary model (Guion, 2000). That is, a test may be considered fair as long as it *does not exhibit bias against* the minority subgroup. This idea is especially apparent in much of the literature discussing the "fairness" of cognitive ability tests, as evidenced by statements such as

"The general acceptance of the Cleary (1968) model…" and "acceptance of the Cleary model as 'superior'…" (Chung-Yan & Cronshaw, 2002). Further, according to Schmidt and Hunter (1982) "cognitive tests are fair to minority group members (specifically blacks and hispanics) under this model of test fairness." (pp. 601 – 602). Thus, it seems to be the consensus that, if a test exhibits bias, that test is considered fair if long as it is biased in favor of the minority.

Finally, both of the models of test bias/fairness discussed above assumed that the regression lines for the two population subgroups are parallel. However, it may be possible that this is not the case. That is, the slopes of the regression lines may differ significantly. This raises the issue of the possibility of differential validity, i.e., the possibility that a test predicts differentially for the two subgroups. However, previous research has demonstrated that, if the slopes of the regression lines do differ for the two population subgroups, that difference is caused by chance or by statistical artifacts (Hunter, Schmidt, & Hunter, 1979). A test is considered unfair if it predicts performance differently (significantly) for each of the subgroups for which it is administered. Differential prediction can be demonstrated by a finding of a significant effect of the interaction of subgroup membership (in this case, race) and test score on task performance.

In sum, as demonstrated by the discussion above, test bias/fairness is an important issue in selection testing. Thus it is a research question that must be addressed when examining the effectiveness of any new test. Therefore, in the current study, I examined the research question that addresses the possibility of bias and the issue of fairness associated with each of the knowledge tests, as well as the cognitive ability test.

Research Q1: Is the cognitive ability test, the basic knowledge test, and/or the understanding test biased against any subgroup compared to the other subgroup?

Method

Participants

Participants ($N = 320$) were students at a medium-sized mid-western university with an open-enrollment policy. All students participated in exchange for extra credit points. Of these students, I obtained 145 (45%) from a pool of undergraduate students enrolled in introductory psychology classes. However, this student pool did not include an adequate number of African Americans to meet the study requirements. Therefore, I recruited the remaining 55% of the participants from outside of the subject pool. Thus, 96 (30%) of the participants were recruited from advanced psychology classes and 81 (25%) from courses other than psychology. Table 1 shows the breakdown of participants recruited from specific courses.

Table 1

Breakdown of Participants by Course Type

Number of Students	Course Type
145	Introductory Psychology
96	Advanced Psychology
6	Anthropology
6	Applied Behavioral Science
18	Regional Studies
34	Sociology
15	Women's Studies
320	

Furthermore, this sample included students from a wide range of university majors. Table 2 lists the breakdown of participants by college major.

Table 2

Breakdown of Participants by College Major

College Major	Number of Participants
Applied Behavioral Science	5
Biology	2
Bio-Medical Engineering	1
Business	22
Communications	7
Computer Science	2
Criminal Justice	3
Dance	1
Education	19
Engineering	6
Exercise Biology	2
History	1
Mass Communications	1
Music Education	1
Organizational Leadership	1
Psychology	118
Social Work	1
Sociology	10
Nursing	11
Urban Affairs	1
Undecided	2
Other	103
	320

Thus, the final sample consisted of 320 students, ranging in age from 17 to 53 years. Of these, 123 were African Americans (19 males, 104 females); 179 were White Americans (64 males, 115 females), 4 were Asian Americans (1 male, 3 females), 9 were Hispanic Americans (4 males, 5 females), and 5 listed their race as "Other" (2 males, 3 females).

Task Description

Participants performed an adaptation of an existing computer-based truck dispatching simulation task (Steele-Johnson & Perlow, 1989). This task required participants to receive, process, and ship orders of military parts and supplies to various areas within an Area of Responsibility (A.O.R.). Eight rules constrained the processing of these orders. Rules included information about identifying order types, order delivery schedules, truck capacity restrictions, and delivery area restrictions. Running task time was identified as the time spanning one working week. That is, the task time began at 9:00 a.m. on Monday and ended at 5:00 p.m. Friday. Each hour interval in the task was equivalent to 15 seconds in real time. Each day in the task corresponded to two minutes in real time. Each task trial lasted 10 minutes. All participants received task instructions prior to attempting the task.

Test Types

A primary purpose of the current study was to examine the effects of three types of tests on perceptions and performance: tests of cognitive ability, basic knowledge, and understanding.

Cognitive ability. I assessed cognitive ability using the *Wonderlic* Personnel Test (Wonderlic, C., 2002). The measure is a 12-minute, speeded test, with a total of 50 items. This measure assesses verbal, mathematical, and analytical ability. I calculated the test score as the sum of correct responses. The reported test-retest reliability of this measure ranges from .82 to .94.

Basic knowledge. I assessed basic knowledge using a measure designed for this study (see Appendix A). I selected the items included in the current test from an item bank that was developed for use in the author's master's thesis. That basic knowledge test employed existing industrial/organizational test-development procedures (see Palumbo, 2004, for a discussion of test design methods). I designed the test to measure an individual's knowledge of the declarative information (i.e., rules) and procedural information (i.e., keystroke use) that governs the task. I used a multiple-choice framework with four response options for this 30-item, paper-and-pencil measure and calculated the test score as the sum of the correct responses. Sample questions: "What is accomplished by pressing CNTRL-P?" [Answer: "You can ship the pickup order(s)."] and "What key or keys allow you to ship a pickup order?" (Answer: "CNTRL-P.") These questions assessed only declarative and procedural task information. The internal reliability coefficient for this test in this study was .84.

Understanding. I assessed understanding using a paper-and-pencil measure designed for this study (see Appendix B). I selected the items included in the current test from an item bank that was developed for the author's master's thesis (See Palumbo, 2004, for a discussion of test design methods). That understanding test employed a cognitively-oriented approach to test development. I designed the test to measure the qualitative components of an individual's overall knowledge and understanding of task requirements, procedures, and their appropriate applications for successful task completion. I used a multiple-choice framework with five response options for this 30-item, paper-and-pencil measure and calculated the test score as the sum of the correct responses. Sample questions: "If you attempt to ship a truck that is overloaded by 5 units:" (Answer: "The truck will go out, and a penalty is assessed.") and "If you attempt to ship a truck that is loaded with 42 or more items:" (Answer: "The truck will not go out, and a penalty is assessed."). These sample questions required the individual to understand that there was a predetermined maximum truck capacity, that they had exceeded truck capacity by a specific amount, and that consequences existed for attempting to ship that specific truck. The internal reliability coefficient for this test in this study was .80.

Test Order

To examine the possibility that exposure to one knowledge test influences perceptions of and performance on the next, I implemented two test order conditions. More specifically, it may be the case that individuals' perceptions of an initial test might influence their perceptions of a second test. Additionally, it may be the case that individuals acquire knowledge (learn) from the information included in the first test, beyond their level of task-related information acquired during the task. This additional information, then, may enhance performance on the second knowledge test. Therefore, I divided participants into two groups such that one half received the basic knowledge test, followed by the understanding test; the second group received the knowledge tests in the opposite order.

Measures

Test perceptions. I assessed test perceptions using a 13-item, paper-and-pencil measure (see Appendix C), which was designed for this study to capture the individual's perceptions of three distinct constructs: 1) perceptions of the tests as measures of intelligence and knowledge, 2) test-task relatedness (face validity), and 3) test-taking motivation. Participants rated the degree to which they agreed with each question on a 7-point Likert-type scale with 1 representing "Strongly Disagree" and 7 representing "Strongly Agree".

I assessed *perceptions of the tests as measures of intelligence and knowledge* using an adaptation of Steele and Aronson's (1995) manipulation check for stereotype threat inductions. Their single-item measure began with an open-ended statement, followed by three possible response options. Participants responded by choosing one of the three possible options. The item in their original measure was: "The purpose of this experiment was to: (a) provide a genuine test of my abilities in order to examine personal factors involved in verbal ability; (2) provide a challenging test to examine factors involved in solving verbal problems; (c) present you with unfamiliar verbal problems to measure verbal learning." (Steele & Aronson, 1995, p. 800).

The measure I designed for this study included two questions (Subscale 1), designed to assess the individual's perceptions of each test as a measure of intelligence and as a measure of job knowledge. The questions were: "1. I believe that the test I just completed was designed to measure of my general intelligence."; and "2. I believe that the test I just completed was designed to measure my knowledge of a particular task."

I assessed *test relatedness perceptions* using an adaptation of a 4-item scale designed by Chan, Schmitt, Sacco, and DeShon, (1998). These authors designed the items to assess the individual's perception of the relatedness of the test to the job. In Chan et al.'s study, participants responded on a 5-point, Likert-type scale with 1 representing "Strongly Disagree" and 5 Representing "Strongly Agree". A sample item from the test relatedness scale is: "I do not understand what the test had to do with this job." Thus, Subscale 2 of the test perceptions measure used in this study, Items 3 - 6, consisted of Chan, Schmitt, Sacco, and DeShon's (1998) test relatedness items. Chan et al. (1998) reported a reliability coefficient of .74 for this scale. In the current study, the internal reliability coefficient was .75. These items were administered to enable the possible examination of alternative explanations for the observed results.

I assessed *test-taking motivation* using a 7-item scale designed by Chan, Schmitt, DeShon, Clause, and Delbridge, (1997). These authors designed the items to assess the individual's perception of their level of test-taking motivation. In Chan et al.'s study, participants responded on a 5-point, Likert-type scale with 1 representing "Strongly Disagree" and 5 Representing "Strongly Agree". A sample item from the test-taking motivation scale is: "Doing well on this test was important to me." Thus, Subscale 3 of the test perceptions measure used in this study, Items 7 - 13, consisted of Chan, Schmitt, DeShon, Clause, and Delbridge, (1997) test-taking motivation items. Chan et al. (1998) reported a reliability coefficient of .86 for this scale. In the current study, the internal reliability coefficient was .91. These items were administered to enable the possible examination of alternative explanations for the observed results.

Task performance. I operationalized task performance as participants' performance scores on each trial, as calculated by the computer. Participants received 5 points for each unit of office equipment correctly shipped and lost 10 points for each rule violation.

Demographics. I collected demographic information using a paper-and-pencil measure designed for this study (see Appendix D). Demographic items included age, race, gender, college ranking (freshman, sophomore, junior, senior), and college major.

General perceptions. I assessed general perceptions about the fixedness (or malleability) of both intelligence and job knowledge using a paper-and-pencil measure designed for this study. This measure included two questions: "1. I believe that my intelligence is a fixed, unchanging characteristic."; "2. I believe that my knowledge about a particular task can be changed." Participants rated the degree to which they agree with each question on a 7-point Likert-type scale with 1 representing "Strongly Disagree" and 7 representing "Strongly Agree".

Procedure

Participants were assigned to one of two test order conditions such that half of the participants received the basic knowledge test, followed by the understanding test; the second half received the knowledge tests in the opposite order. Figure 1 shows the order of administration of experimental procedures for both testing conditions.

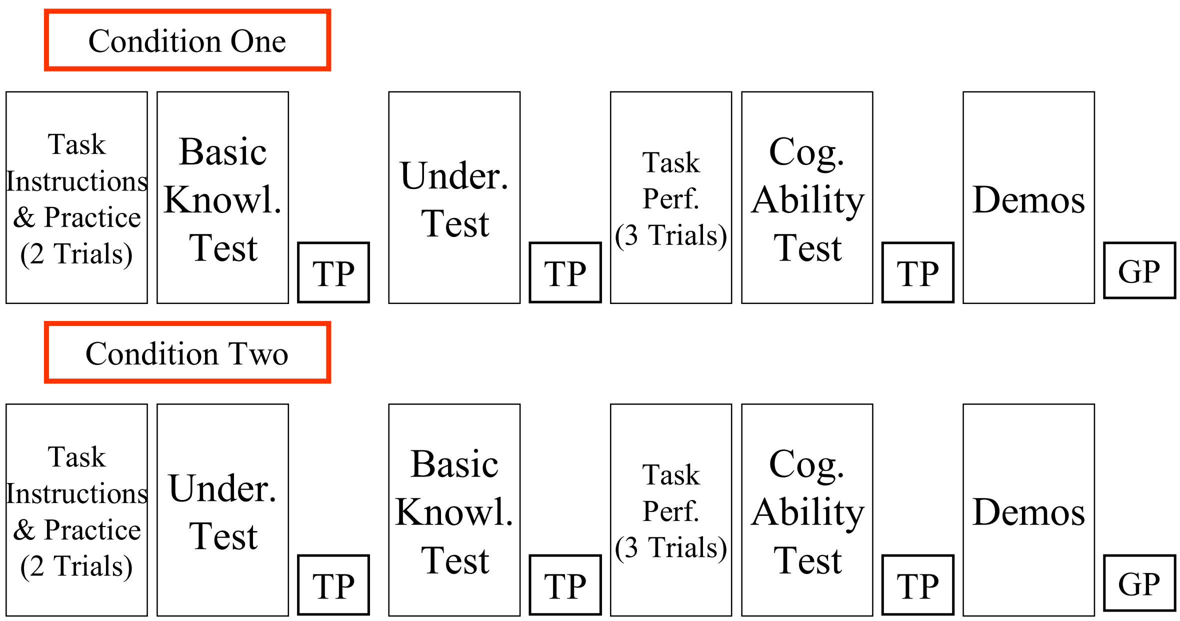

Figure 1. Model denoting order of administration of experimental procedures.

Each participant received task instructions and performed two 10-minute practice trials (Trials 1 and 2). Next, participants completed the written measures of basic knowledge and task understanding, with the order depending on condition. Immediately following the first job knowledge test, participants completed the measure of test perceptions (TP) for that test. Then, participants completed the second job knowledge test followed immediately by another measure of test perceptions (TP). Next, participants completed three more 10-minute trials of the dispatching task (Trials 3, 4, and 5). After completing the task trials, each participant completed the measure of cognitive ability. Administering the cognitive ability test after the job knowledge tests eliminated the possibility that any stereotype threat effect triggered by the cognitive ability test would affect performance on the job knowledge tests. Next, participants completed a test perceptions (TP) measure relative to the cognitive ability test. Finally, participants completed the demographics questionnaire and the general perceptions (GP) measure at the end of the experimental session. Note: Providing demographic information only at the end of the session eliminated the possibility of making race salient to the participants prior to their completion of knowledge or ability measures.

Results

Sample Characteristics

Participants (N = 320) were run in one of two conditions which differed only in the order of administration of the two job knowledge tests. In Condition 1, the basic knowledge test preceded the understanding test. In Condition 2 the test order was reversed. Out of a possible 1600 task performance scores (320 participants X 5 scores per person), a total of 37 (2.3%) missing values were observed. These missing task performance data were replaced by calculating a regression line for each individual and estimating the missing value according to the regression. In sum, a total of 13 (0.8%) scores were estimated for performance on the two practice trials (8 scores for T-1; 5 for T-2). Additionally, 24 (1.5%) scores were estimated for performance on the three task trials following knowledge test administration: 10, 5, and 9 scores for Trials 3, 4 and 5, respectively.

Further, the data were examined for the presence of possible outliers and/or response bias. Using a criterion cutoff of five (5) standard deviations (SD) from the mean for performance on all tests and the dispatching task, no outlying values were observed. All values fell within three (3) SDs of the mean value. Therefore, all individual data were available for use in the analyses.

In order to be compare even numbers of African American participants to Whites, the sample had to be reduced. I accomplished this reduction by removing, initially, all (18) participants who listed their race as "Asian American" (4 participants), "Hispanic American" (9 participants), or "Other" (5 participants). Then, using a random number generator, I selected participants to remove on the basis of race. Of the resulting sample (n = 246), 104 African Americans were female, 19 were male; 62 Whites were female, 61 were male. This sample was used in all analyses.

Means, Standard Deviations, and Intercorrelations between Study Variables

Means, standard deviations, and intercorrelations between Study 1 variables are shown in Table 3.

Table 3

Means, Standard Deviations, and Correlations of Study Variables

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Race	-----	-----	--															
2 TPCAT1	5.99	1.51	.10	--														
3 TPCAT2	3.11	2.08	-.19**	-.05	--													
4 TPBas1	3.33	1.88	-.19**	.24***	.36†	--												
5 TPBas2	5.28	1.79	-.01	.21***	-.14*	.07	--											
6 TPUnd1	3.35	1.93	-.19**	.23***	.37†	.73†	.08	--										
7 TPUnd2	5.35	1.71	-.06	.29†	-.04	.12	.55†	.11	--									
8 GenPerc1	2.30	1.58	-.07	-.14*	.09	.12	-.04	.02	-.10	--								
9 GenPerc2	6.46	1.00	.15*	.17**	-.16*	-.12	.19**	-.09	.20**	-.30†	--							
10 WonTot	21.34	5.87	.48†	.10	-.32†	-.29†	.00	-.29†	-.03	-.06	.21***	--						
11 BasTot	18.84	5.77	.34†	.11	-.26†	-.10	.20**	-.03	.11	-.11	.21***	.54†	--					
12 UndTot	13.18	5.39	.42†	.12	-.29†	-.13*	.11	-.06	.08	-.13*	.24***	.58†	.77†	--				
13 PerfT3	648.84	819.41	.35†	.07	-.14*	-.11	.04	.01	.00	-.11	.07	.37†	.56†	.59†	--			
14 PerfT4	994.92	898.33	.39†	.07	-.12	-.06	.06	.05	.05	-.10	.08	.35†	.53†	.55†	.81†	--		
15 PerfT5	1210.07	893.50	.37†	.02	-.12	-.06	.07	.04	.01	-.11	.03	.29†	.49†	.50†	.75†	.89†	--	
16 PerfAvg	951.27	817.04	.39†	.05	-.13*	-.08	.06	.03	.02	-.11	.06	.36†	.56†	.58†	.91†	.96†	.94†	--

* $p < .05$, ** $p < .01$, *** $p < .001$, † $p < .0001$.

Key:

TPCAT1	Perceptions that the Cognitive Ability Test (CAT) is a measure of general intelligence
TPCAT2	Perceptions that the Cognitive Ability Test (CAT) is a measure of job/task knowledge
TPBas1	Perceptions that the Basic Knowledge Test is a measure of general intelligence
TPBas2	Perceptions that the Basic Knowledge Test is a measure of job/task knowledge
TPUnd1	Perceptions that the Understanding Test is a measure of general intelligence
TPUnd2	Perceptions that the Understanding Test is a measure of job/task knowledge
GenPerc1	Perceptions of Intelligence as fixed, unchangable
GenPerc2	Perceptions of Job Knowledge as Malleable
WonTot	Total Score on the Wonderlic Cognitive Ability Test
BasTot	Total Score on the Basic (Job Knowledge) Test
UndTot	Total Score on the Understanding (Job Knowledge) Test

PerfT3 - PerfT5 refers to Task Performance in Trials 3 through 5

Initial Group Differences

To check for initial group differences, I compared the means for the variables within each condition (i.e., Test Order). As stated earlier, in both groups, 50% of the participants were African American and 50% were White. In both test order groups, 67.5% were female and 32.5% were male. Also, the breakdown for each race and gender were similarly dispersed across both testing conditions [Condition 1: 30 white males, 30 white females, 9 black males, 51 black females; Condition 2: 31 white males, 32 white females, 10 black males, 53 black females]. Further, although the age ranges for the two groups differed slightly (i,e., participants in Test Order 1 ranged in age from 18 to 53 and from 17 to 45 in Test Order 2), the median age within each condition was 19.5 years and the mean age was 22 years. Further, the cognitive ability scores ranged from 10 to 36 in Condition 1 [$M = 21.10$, $SD = 5.59$] and from 7 to 35 in Condition 2 [$M = 21.57$, $SD = 6.13$]. Results indicated no significant difference in cognitive ability in the two conditions [$F(1, 245) = 0.40$, $p = .52$]. Therefore, I concluded that minimal differences existed between the two groups and that race, gender, and cognitive ability were evenly distributed across both test order conditions.

Effects of Test Order

First, I examined the effects of test order on perceptions and performance to determine whether test order influenced perceptions of tests as measures of intelligence and knowledge, performance on the three tests, or performance on the task.

Effects of condition (test order) on perceptions of tests as measures of intelligence and knowledge. To examine whether test order influenced perceptions of tests as measures of intelligence and knowledge, I performed two anovas with test type as the within subject factor. Specifically, I examined the effect of test order on the perception of the test as a measure of intelligence for the two knowledge tests. Results demonstrated a significant effect for condition with participants reporting stronger perceptions of the tests as measuring intelligence when the basic knowledge test was administered first [$F(1, 244) = 8.99$, $p = .003$]. Then I examined the effect of test order on the perception of the test as a measure of knowledge for the two tests. Resulted indicated a significant Condition X Test Type interaction effect [Wilks' Lambda = .95, $F(1, 244) = 12.60$, $p = .0005$]. Participants reported a stronger perception of the basic knowledge as a measure of knowledge when the understanding test was administered first [$F(1, 244) = 12.72$, $p =$

.0004], but test order had no effect on perceptions of the understanding test as a measure of knowledge [$F(1, 244) = 1.34$, $p = .25$]. This effect did not play a role in tests of hypotheses and only had implications for exploratory analyses with perceptions as an outcome variable.

Effects of condition (test order) on test performance. To examine whether test order influenced performance on the two knowledge tests, I performed two anovas. Specifically, I examined the possible existence of 'learning from the test' by examining the effects of test order on test performance for each of the two knowledge tests. Results demonstrated no significant effect for order of test administration on scores on the basic knowledge test [$F(1, 245) = 0.01$, $p = .92$] or scores on the test of understanding [$F(1, 245) = 0.71$, $p = .40$].

Effects of condition (test order) on task performance. Further analyses demonstrated that the order of knowledge test administration had no significant effect on task performance in any of the post-test task performance trials: $F(1, 245) = 0.01$, $p = .91$, $F(1, 245) = 1.42$, $p = .23$, and $F(1, 245) = 1.84$, $p = .18$, for Task Trials 3 - 5, respectively. Therefore, the order of test administration was excluded from tests of hypotheses.

Effects of test type on perceptions. To examine whether test type (cognitive ability, basic knowledge, understanding) was effective in inducing perceptions of tests as measures of intelligence and knowledge, I performed a repeated measures anova with test type as the within subject factor. I conducted the analysis twice, once for the perception of the test as a measure of intelligence and once for the perception of the test as a measure of knowledge. Results indicated a significant effect for test type on perceptions of the test as a measure of intelligence [Wilks' Lambda = 0.36, $F(2, 244) = 211.12$, $p < .0001$]. Additionally, results indicated a significant effect for test type on perceptions of the test as a measures of (job) knowledge [Wilks' Lambda = 0.59, $F(2, 244) = 82.44$, $p < .0001$].

Further, I tested contrasts for each perception item, comparing perceptions associated with the ability test versus perceptions associated with the basic knowledge or understanding tests. Results indicated a significant difference in the perceptions of the tests as *measures of intelligence* [$F(2, 490) = 312.05$, $p < .0001$]. To examine these effects further, I contrasted the perceptions of the cognitive ability test with each of the job knowledge tests. Specifically, I calculated the means for perceptions for the three tests. Then, using t-tests (two-tailed), I compared the means for each of the two knowledge test against the mean of the cognitive ability test, and against each other. This procedure was completed twice–once for comparing the perceptions of the tests as measures of intelligence and again for comparison of perceptions of the tests as measures of knowledge. Although these tests were nonorthogonal, they provided me with a more complete description of my data.

When the cognitive ability test was compared with the basic knowledge test, results indicated a significant difference in test perceptions. When the cognitive ability test was compared with the understanding test, results indicated a significant difference in test perceptions. Both knowledge tests were perceived as different from the cognitive ability test, such that participants reported a mean of 5.99 [$SD = 1.51$] for perceptions that the cognitive ability test measured intelligence as compared to a mean of 3.33 [$SD = 1.88$, $t(245) = 19.68$] and a mean of 3.35 [$SD = 1.92$, $t(245) = 19.23$], respectively, for the basic knowledge test and the test of understanding, both *p*s < .0001. Thus, the cognitive ability test was perceived more strongly as a measure of intelligence than either of the job knowledge tests. Furthermore, the contrast between each of the two knowledge tests yielded no significant difference [$t(245) = -0.27$, $p = .78$]. Thus both knowledge tests were perceived as significantly different from the cognitive ability test but not significantly different from each other.

Also, results indicated a significant difference in the perceptions of the tests as *measures of knowledge* [$F(2, 490) = 125.92$, $p < .0001$]. To examine these effects further, I contrasted the perceptions of the cognitive ability test with each of the job knowledge tests. Specifically, using t-tests (two-tailed), I compared the mean for perceptions of the cognitive ability tests as a measure of knowledge against the means of perceptions of each of the two knowledge tests as measures of knowledge.

When the cognitive ability test was compared with the basic knowledge test, results indicated a significant difference in test perceptions. When the cognitive ability test was compared with the understanding test, results indicated a significant difference in test perceptions. That is, results indicated a significant difference in the way that the cognitive ability test was viewed relative to both knowledge tests [$t(245) = -11.59$ and -12.74, respectively, for the basic knowledge test and the test of understanding, both *p*s < .0001]. The reported means for perceptions of the three tests as measures of knowledge were $M = 3.11$, $SD = 2.08$; $M = 5.28$, $SD = 1.79$; and $M = 5.35$, $SD = 1.71$; respectively, for perceptions of the cognitive ability test, the basic knowledge test, and the understanding test as a measure of knowledge. Thus, the two knowledge tests were perceived more strongly as measures of knowledge than was the cognitive ability test.

Finally, the contrast between each of the two knowledge tests yielded no significant difference [$t(245) = -0.27$, $p = .54$]. Thus both knowledge tests were perceived as significantly different from the cognitive ability test but not significantly different from each other.

The Role of Race

An overarching theme developed in the introduction is that, for any given test, race influences test perceptions, which in turn affect test performance (i.e., test scores), which ultimately influences task performance. Researchers have discussed some of these links conceptually (e.g., Steele & Aronson, 1995; Steele, 1997) and have provided empirical evidence supporting other links (e.g., Schmidt & Hunter, 1998). In the current study, I examined some of these links as stated hypotheses; however, I conducted additional preliminary analyses examining other links in the conceptual model. Specifically, I conducted preliminary analyses examining the effects of 1) race on test perceptions, 2) race on general perceptions, 3) race on task performance, 4) test perceptions on test performance, and 5) test perceptions on task performance.

Race effects on test perceptions. To examine the role of race on test perceptions, I performed a series of anovas, with race as the predictor and test perceptions as the outcome. Specifically, I examined the relationship between race and each test perception, individually, relative to each test type. That is, I examined the relationship between race and the perception of the test as measuring intelligence for each of the three (cognitive ability, basic knowledge, and understanding) tests. Also, I examined the relationship between race and the perception of the test as measuring knowledge for each of the three tests.

Results indicated no significant effect for race on the perception of the *cognitive ability test as a measure of intelligence* [$F(1, 244) = 2.31$, $p = 0.129$]. Participants reported similar perceptions [$M = 5.85$, $SD = 1.67$; $M = 6.14$, $SD = 1.33$, for Blacks and Whites, respectively]. However, the effect of race on the perception of the *cognitive ability test as a measure of knowledge* was significant [$F(1, 244) = 8.92$, $p < .01$]. A comparison of means for the two groups demonstrated that Blacks perceived the cognitive ability test as a measure of job/task knowledge more strongly than did Whites [Blacks: $M = 3.50$, $SD = 2.23$; Whites: $M = 2.72$, $SD = 1.85$].

Additionally, results indicated a significant effect for race on the perception of the *basic knowledge test as a measure of intelligence* [$F(1, 244) = 8.94$, $p < .01$]. A comparison of means for the two groups demonstrated that Blacks perceived the basic knowledge test as a measure of intelligence more strongly than did Whites [Blacks: $M = 3.68$, $SD = 1.92$; Whites: $M = 2.98$, $SD = 1.79$]. However, the effect of race on the perception of the *basic knowledge test as a measure of knowledge* was not significant [$F(1, 244) = 0.01$, $p = .92$; Blacks: $M = 5.29$, $SD = 1.83$; Whites: $M = 5.27$, $SD = 1.75$].

Finally, a pattern similar to that observed for the basic knowledge test emerged for perceptions of the understanding test. Specifically, results indicated a significant effect for race on the perception of the *understanding test as a measure of intelligence* [$F(1, 244) = 9.39$, $p < .01$]. A comparison of means for the two groups demonstrated that Blacks perceived the understanding test as a measure of intelligence more strongly than did Whites [Blacks: $M = 3.72$, SD =1.99; Whites: M =2.98, SD = 1.79]. However, the effect of race on the perception of the *understanding test as a measure of knowledge* was not significant [$F(1, 244) = 1.01$, $p = 0.32$ [Blacks: $M = 5.46$, $SD = 1.64$; Whites: M =

5.24, *SD* = 1.78]. Thus race had a significant effect in the perceptions associated with the understanding test.

Race effects on general perceptions (beliefs) about intelligence and job knowledge. To examine the role of race in general perceptions that intelligence is fixed and/or that job knowledge is malleable, I performed anovas, with race as the predictor and general perceptions as the outcome. Specifically, I assessed the effects of race on each type of general perception (intelligence as fixed, knowledge as malleable). Results indicated no significant effect of race on the perception that *intelligence is fixed* [$F(1, 244) = 1.19$, $p = 0.28$; [Blacks: $M = 2.41$, $SD = 1.68$; Whites: $M = 2.19$, $SD = 1.46$]. However, results indicated a significant effect of race on the perception (belief) that *job knowledge is malleable* [$F(1, 244) = 5.72$, $p < .05$]. A comparison of means indicated that Blacks perceived job knowledge as less malleable than did Whites [Blacks: $M = 6.31$, $SD = 1.20$; Whites: $M = 6.60$, $SD = 0.72$].

The effects of race on task performance. To examine the effects of race on task performance, I performed anovas with repeated measures. Performance in the three post-test task trials (T3 – T5) was the within subject factor. Results indicated a significant effect for Trial [Wilks' Lambda = 0.85, $F(2, 243) = 21.91$, $p < .0001$]. Performance improved across trials (M = 648.84, 994.92, and 1210.07 for Trials 3, 4, and 5, respectively). Additionally, results indicated a significant effect for race [$F(1, 244) = 45.01$, $p < .0001$]. Blacks performed lower than Whites [Blacks' average performance across trials: 624.49, SD =735.32; Whites' average performance: 1273.06, SD = 768.82]. There were no other significant effects.

The role of test perceptions in test performance. To examine the effects of test perceptions on test performance, I performed a series of regressions in which test performance (test score) was regressed on each related perception (predictor). Specifically, I regressed the score on the cognitive ability test on each perception relative to that test. I repeated the procedure for each perception for each of the knowledge tests. Results indicated no significant effect for perceptions of the *cognitive ability test* as a measure of intelligence on the score on the cognitive ability test [$F(1, 244) = 2.65$, $p = 0.105$, $\beta = 0.104$]. However, results indicated a significant effect for perceptions of the cognitive ability test as a measure of knowledge [$F(1, 244) = 28.60$, $p < .0001$, $\beta = -0.32$]. Thus, perceiving the cognitive ability as a measure of knowledge had a significant ***detrimental*** effect on score on the cognitive ability test.

Next, I examined the effects of test perceptions on test score relative to the *basic knowledge test*. First, I regressed test performance on the perception of the basic test as a measure of intelligence. Results indicated no significant effect [$F(1, 244) = 2.36$, $p = 0.13$, $\beta = -0.098$]. Thus, perceiving the basic knowledge test as a measure of intelligence had no significant effect on test score. However, results indicated a significant effect for perceptions of the basic test was perceived as a measure of knowledge [$F(1, 244) = 9.68$, $p = 0.002$, $\beta = 0.195$]. Thus, perceiving the basic test as measuring knowledge had a significant, positive effect on test score.

Finally, I examined the effects of test perceptions on test score relative to the *understanding test*. First, I regressed test performance on the perception of the understanding test as a measure of intelligence. Results indicated no significant effect [$F(1, 244) = 0.82$, $p = 0.36$, $\beta = -0.058$]. Further, results indicated no significant effect for the perception of the understanding test as a measure of knowledge [$F(1, 244) = 1.57$,

$p = 0.21$, $\beta = 0.079$]. Thus, neither perception relative to the test of understanding had a significant on the understanding test score.

The role of test perceptions in task performance. To examine the effects of test perceptions on task performance, I performed a series of repeated measures regressions with task performance (score) in Trials 3 - 5 as the within subject factor. First, I examined the effect of the perception of the test as a measure of intelligence on performance for the *cognitive ability test*. Results indicated a significant effect for Trial [Wilks' Lambda = 0.91, $F(2, 243) = 11.39$, $p < .0001$]. Performance improved across trials ($M = 648.84$, $SD = 819.41$; *994*.92, $SD = 898.33$; and 1210.07, $SD = 893.50$ for Trials 3, 4, and 5, respectively). There were no other significant effects for the perception of the *cognitive ability test* as a measure of intelligence. Next, I examined the effect of the perception of the test as a measure of knowledge on performance. Results indicated a significant effect of trial [Wilks' Lambda = 0.80, $F(2, 243) = 30.68$, $p < .0001$]. These means are reported above. Further, results indicated a main effect for the perception of the cognitive ability test as a measure of knowledge [$F(1, 244) = 4.44$, $p < .05$, $\beta = -0.134$]. Thus, perceptions of the cognitive ability test *as a measure of knowledge* had a ***detrimental*** effect on average task performance. There were no other significant effects.

Next, task performance in Trials 3 - 5 was regressed on each perception for the *basic knowledge test*. In examining the effect of the perception of the test as a measure of intelligence, results indicated a significant effect for Trial [Wilks' Lambda = 0.88, $F(2, 243) = 20.26$, $p < .0001$]. Performance improved across trials ($M = 648.84$, $SD = 819.41$; *994*.92, $SD = 898.33$; and 1210.07, $SD = 893.50$ for Trials 3, 4, and 5, respectively).

There were no other significant effects. Then I examined the effect of the perception of the basic knowledge test as a measure of knowledge. Results indicated a significant effect for Trial [Wilks' Lambda = 0.94, $F(2, 243) = 7.58$, $p < .001$]. Performance improved across trials (see means noted above). There were no other significant effects.

Finally, I regressed task performance in Trials 3 - 5 on each perception for the *understanding test*. In examining the effect of the perception of the test as a measure of intelligence, results indicated a significant effect for Trial [Wilks' Lambda = 0.85, $F(2, 243) = 21.50$, $p < .0001$]. Performance improved across trials ($M = 648.84$, $SD = 819.41$; *994*.92, $SD = 898.33$; and 1210.07, $SD = 893.50$ for Trials 3, 4, and 5, respectively). There were no other significant effects. Then I examined the effect of the perception of the understanding test as a measure of knowledge. Results indicated a significant effect for Trial [Wilks' Lambda = 0.92, $F(2, 243) = 10.06$, $p < .001$]. Performance improved across trials (see means reported above). There were no other significant effects. Thus, neither perception relative to the test of understanding had any significant on task performance.

Size of Black-White Mean Differences in Scores on the Ability, Basic Knowledge, and Understanding Tests (Hypothesis 1)

To examine the effects of test type on Black-White mean differences in test scores (Hypotheses 1a, 1b, and 1c), I first computed the means for each test score within the subgroups. Then, I computed t-tests comparing test scores for blacks versus whites on each of the three tests. Finally, I examined whether the effect sizes observed for the three t-tests are significantly different.

Specifically, to compare effect sizes I followed the approach suggested by

Valentine and Cooper (2003). I assessed (a) the statistical significance of the effects, (b) the practical significance based on raw mean differences, and (c) the relative size of the effects based on standardized estimates of effect size. I accomplished (a) as described in the preceding paragraph; (b) by reporting raw mean differences; and (c) by converting the reported t values for the groups into standardized mean difference statistics, referred to as d (Cohen, 1988), for direct comparison. Beyond practical interpretation, using Cohen's (1988) established benchmarks for effect size estimates (i.e., small if $d = .20$, medium if $d = .50$, and large if $d = .80$), we can assert whether the effect size for race differences is different in the three tests.

Size of Black-White mean differences on the cognitive ability test. The mean scores on the cognitive ability tests were 18.54, ($SD = 4.86$) and 24.14, ($SD = 5.44$) for Blacks and Whites, respectively. Thus, Blacks scored ***more*** than one SD lower than Whites on the cognitive ability test, thereby providing support for Hypothesis 1a (H_{1a}). Results of the t-test indicated a significant difference in test scores [$t(122) = 8.26$, $p < .0001$. The raw mean difference was 5.6 lower for Blacks than Whites. The standardized difference (Cohen's d) was 1.03 such that Blacks scored 1.03 SD lower than Whites. Thus, there was a large effect for the difference between Blacks and Whites on the cognitive ability test.

Size of Black-White mean differences on the basic knowledge test. The mean scores on the basic knowledge tests were 16.88, ($SD = 5.70$) and 20.79, ($SD = 5.16$) for Blacks and Whites, respectively. Thus, Blacks scored ***less*** than one SD lower than Whites on the basic knowledge test. Results of the t-test indicated a significant difference in test scores [$t(122) = 5.86$, $p < .0001$. The raw mean difference was 3.91

lower for Blacks than Whites. The standardized difference (Cohen's *d*) was 0.75 such that Blacks scored 3/4ths of one SD lower than Whites. However, similar to the effect size found for the cognitive ability test, these results indicate a large effect for the difference between Blacks and Whites on the basic knowledge test. Thus, Hypothesis 1b (H_{1b}) was not supported.

Size of Black-White mean differences on the understanding test. The mean scores on the understanding tests were 10.89, (SD = 4.99) and 15.46, (SD) = 4.79 for Blacks and Whites, respectively. Thus, Blacks scored ***less*** than one SD lower than Whites on the test of understanding. Results of the t-test indicated a significant difference in test scores [$t(122) = 37.05$, $p < .0001$. The raw mean difference was 4.57 lower for Blacks than Whites. The standardized difference (Cohen's *d*) was 0.95 such that Blacks scored 95% of one SD lower than Whites. However, these results indicate also a large effect for the difference between Blacks and Whites on the understanding test. Thus, Hypothesis 1c (H_{1c}) was not supported.

The Role of Cognitive Ability and Knowledge as Predictors of Performance (Hypothesis 2)

To examine the effectiveness of cognitive ability, basic knowledge, and understanding for predicting performance, I performed a series of three regression analyses with repeated measures. Specifically, I regressed performance in the three post-test task trials (T3 – T5) on the cognitive ability, basic knowledge, and understanding test scores, individually. Each regression allowed me to determine the amount of variance in performance (R^2) that was associated with each of the predictors. I would have support for my predictions (H_{2a}, H_{2b}, H_{2c}) if the amount of variance accounted for is largest for understanding, smaller for basic knowledge, and smallest for cognitive ability.

Cognitive ability effects on task performance. The first repeated measures regression analysis examined the effect of cognitive ability on task performance in Trials 3 - 5. Results indicated a significant effect for Trial [Wilks' Lambda = 0.90, $F(2, 243) = 13.30$, $p < .0001$]. Performance improved across trials (M = 648.84, 994.92, and 1210.07 for Trials 3, 4, and 5, respectively). Additionally, results indicated a significant main effect for cognitive ability [$F(1, 244) = 35.38$, $p < .0001$, $\beta = 0.36$]. There was no significant interaction effect [Trial X Cognitive Ability: Wilks' Lambda = 0.98, $F(2, 243) = 2.06$, $p = 0.13$].

Basic knowledge effects on task performance. The second repeated measures regression analysis examined the effect of basic knowledge on performance in Trials 3 - 5. Results indicated a significant effect for Trial [Wilks' Lambda = 0.90, $F(2, 243) = 12.63$, $p < .0001$]. Performance improved across trials. Additionally, results indicated a significant main effect for basic knowledge [$F(1, 244) = 109.60$, $p < .0001$, $\beta = 0.56$]. There was no significant interaction effect [Trial X Basic knowledge: Wilks' Lambda = 0.99, $F(2, 243) = 1.21$, $p = 0.30$].

Understanding effects on task performance. The third repeated measures regression examined the effect of understanding on performance in Trials 3 - 5. Results indicated a significant effect for Trial [Wilks' Lambda = 0.85, $F(2, 243) = 21.40$, $p < .0001$]. Performance improved across trials. Additionally, results indicated a significant main effect for understanding [$F(1, 244) = 123.27$, $p < .0001$, $\beta =0.58$]. There was no significant interaction effect [Trial X Understanding: Wilks' Lambda = 0.99, $F(2, 243) = 1.79$, $p = 0.17$].

*Test of Hypothesis 2. T*o test Hypotheses 2a, 2b, and 2c directly, I had to compare the variance accounted for by each test. To complete this comparison process, first, I determined the amount of variance (R^2) in performance accounted for by each of the three tests using a series of regression analyses. Regression results indicated that cognitive ability accounted for 13% of the variance in average performance across trials [(R^2) = 0.126], basic knowledge accounted for 31% of the variance [(R^2) = 0.309], and understanding accounted for 34% of the variance in average performance across task trials [(R^2) = 0.336].

Then, to examine the significance of the differences in variance accounted for by the tests, I performed a series of t-test calculations of dependent *r* as described in Cohen and Cohen (1975). I completed this calculation process was complete three times in order to compare: 1) the cognitive ability test to the basic knowledge test, 2) the cognitive ability test to the understanding test, and 3) the basic knowledge test to the understanding test.

Results indicated that the basic knowledge test accounted for significantly more variance than the cognitive ability test [$t(77) = 2.214$, $p < .05$]. Thus Hypothesis 2a was supported. Also, results indicated that the understanding test accounted for significantly more variance than the cognitive ability test [$t(77) = 2.632$, $p < .05$]. Thus Hypothesis 2b was supported. Finally, results indicated that the understanding test failed to account for significantly more variance than the basic knowledge test [$t(77) = 0.673$, $p > .05$]. Thus Hypothesis 2c was not supported.

Knowledge as a Mediator of Ability Effects on Task Performance (Hypothesis 3)

To examine knowledge as a mediator in the cognitive ability - performance relationship (H_{3a}, H_{3b}), I used procedures described by Baron and Kenny (1986) to examine the proposed mediation effects for basic knowledge and understanding.

To demonstrate mediation, the procedures require that preliminary tests demonstrate significant relationships between the variables in each "link" or "pathway" in the proposed model. That is, the predictor variable (cognitive ability) must demonstrate a significant effect on the criterion variable (performance), the predictor variable (cognitive ability) must demonstrate a significant effect on the mediating variable (basic knowledge or understanding), and the mediating variable (basic knowledge or understanding) must significantly affect the criterion variable (performance). Then, to provide evidence of complete mediation, one must demonstrate that the effect of the predictor variable (cognitive ability) on the criterion variable (performance) is reduced to a non-significant level when both the predictor variable (cognitive ability) and mediator (basic knowledge or understanding) variable are entered into the model. However, one can provide evidence of partial mediation if one can demonstrate 1) that the increment or change in the amount of variance (ΔR^2) in the criterion variable accounted for by the mediator variable is not significantly increased by entering the predictor variable into the model, or 2) a significant reduction in the effect of the predictor variable on the criterion variable in the presence of the mediator. In the current study, an increase in the variance accounted for by the predictor variable was tested using a procedure described by Cohen and Cohen (1975) and a reduction in the effect of the predictor variable in the presence of the mediator was assessed using the Sobel test (Baron & Kenny, 1986).

Basic Knowledge as a Mediator of Cognitive Ability Effects on Performance (Hypothesis

3a)

Effect of the predictor (cognitive ability) on the criterion (task performance). As reported above, results demonstrated a significant main effect for cognitive ability on task performance [$F(1, 244) = 35.38$, $p < .0001$, $\beta = 0.36$]. Thus cognitive ability was positively related to task performance. Cognitive ability accounted for 13% of the variance in average performance across trials [(R^2) = 0.126].

Effect of the mediator (basic knowledge) on the criterion (task performance). As reported above, results indicated a significant main effect for basic knowledge [$F(1, 244) = 109.60$, $p < .0001$, $\beta = 0.56$]. Thus, basic knowledge was positively related to task performance. Basic knowledge accounted for 31% of the variance [(R^2) = 0.309] in task performance.

Effect of the predictor (cognitive ability) on the mediator (basic knowledge). I regressed basic knowledge on cognitive ability. Results revealed a significant effect [$F(1, 244) = 99.48$, $p < .001$, $\beta = 0.54$, $R^2 = .289$]. Thus, cognitive ability was positively related to basic knowledge, accounting for 29% of the variance.

Effects of both the predictor (cognitive ability) and the mediator (basic knowledge) on the criterion (task performance). Finally, to examine basic knowledge as a mediator, I entered cognitive ability into the model after controlling for the effects of basic knowledge. Results indicated a main effect for basic knowledge [$F(1, 244) = 66.56$, $\beta = 0.51$, $p < .0001$]. The effect of cognitive ability fell to a non-significant level [$F(1, 244) = 1.58$, $\beta = 0.08$, $p = 0.21$], after controlling for the effect of basic knowledge. The variance in overall task understanding remained at 31% ($R^2 = .314$). That is, no significant incremental variance was accounted for by adding cognitive ability effects (which accounted for 13% of the variance in task performance) into the model. Thus, basic knowledge *completely mediated* the effects of cognitive ability on task performance, thereby failing to provide support for Hypothesis 3a.

Understanding as a Mediator of Cognitive Ability Effects on Performance (Hypothesis 3b)

Effect of the predictor (cognitive ability) on the criterion (task performance). As stated above, cognitive ability had a significant effect on task performance [$F(1, 244) = 35.38$, $p < .0001$, $\beta = 0.36$]. Thus cognitive ability was positively related to task performance, accounting for 13% of the variance in average performance across trials [$(R^2) = 0.126$].

Effect of the mediator (understanding) on the criterion (task performance). As reportd above, understanding had a significant effect on task performance [$F(1, 244) = 123.27$, $p < .0001$, $\beta = 0.58$]. Thus, understanding was positively related to task performance, accounting for 34% of the variance [$(R^2) = 0.336$] in task performance.

Effect of the predictor (cognitive ability) on the mediator (understanding). I regressed understanding on cognitive ability. Results revealed a significant effect [$F(1, 244) = 126.13$, $p < .001$, $\beta = 0.59$, $R^2 = .341$]. Thus, cognitive ability was positively related to understanding, accounting for 34% of the variance.

Effects of both the predictor (cognitive ability) and the mediator (understanding) on the criterion (task performance). Finally, to examine understanding as a mediator, I entered cognitive ability into the model after controlling for the effects of understanding. Results indicated a main effect for understanding [$F(1, 244) = 76.68$, $\beta = 0.56$, $p < .0001$]. The effect of cognitive ability fell to a non-significant level [$F(1, 244) = 0.17$, β

= 0.03, $p = 0.68$] after controlling for the understanding effect. The variance in overall task understanding remained at 34% ($R^2 = .336$). That is, no significant incremental variance was accounted for by adding cognitive ability effects (which accounted for 13% of the variance in task performance) into the model. Thus, understanding *completely mediated* the effects of cognitive ability on task performance, providing support for Hypothesis 3b.

The Role of the Race x Test Performance Interaction in Task Performance (Research Question 1): An Examination of Test Fairness.

To examine the question of test bias/test fairness for all tests (Research Question 1), I performed a series of regressions with repeated measures (i.e., task performance trials) to examine the effects of race and test performance on task performance. The test would be considered unfair if it predicted performance differentially for the two racial subgroups (Blacks and Whites).

Race x Cognitive Ability interaction effects on task performance. I examine first the fairness of the cognitive ability test. Results indicated no significant effects on performance for the between subject effects of race [$F(1, 242) = 0.23$, $p = 0.63$], cognitive ability [Wilks' Lambda = 0.98, $F(1, 242) = 0.95$, $p = 0.33$], or Race X Cognitive Ability [$F(1, 242) = 0.52$, $p = 0.469$]. Also, results indicated no significant effects on performance for the within subject effects for Trial [Wilks' Lambda = 0.98, $F(2, 241) = 2.31$, $p = 0.10$], Trial X Race [Wilks' Lambda = 0.99, $F(2, 241) = 0.798$, $p = 0.45$], Trial X Cognitive Ability [Wilks' Lambda = 0.996, $F(2, 241) = 0.44$, $p = 0.64$], or Trial X Race X Cognitive Ability [Wilks' Lambda = 0.998, $F(2, 241) = 0.29$, $p = 0.75$]. Thus, the cognitive ability test showed no signs of differential prediction.

Race x Basic Knowledge interaction effects on task performance. Next, I examined the fairness of the basic knowledge test. Results indicated significant between subject effects for basic knowledge [$F(1, 242) = 14.53$, $p < .001$] but not for race [$F(1, 242) = .05$, $p = .82$] or Race X Basic Knowledge [$F(1, 242) = 1.08$, $p = .30$]. Thus, basic knowledge significantly affected task performance.

Additionally, results indicated significant within subject effects for Trial X Race [Wilks' Lambda = 0.96, $F(2, 241) = 5.08$, $p < .01$] as well as a significant effect for Trial X Race X Basic Knowledge [Wilks' Lambda = 0.97, $F(2, 241) = 4.09$, $p < .05$], indicating that the basic knowledge test predicted performance differently for the two races in different task trials. In examining this effect, results indicated that the Race X Basic Knowledge interaction effect was significant in Trial 3 [$F(1, 242) = 6.24$, p < .05]. Moreover, follow-up analyses indicated a stronger effect of basic knowledge on task performance for Whites [$F(1, 122) = 52.42$, $p < .0001$, $\beta = .55$] than for Blacks [$F(1, 122) = 31.57$, $p < .0001$, $\beta = .45$]. However, the Race X Basic knowledge effect was non-significant in the later task trials [$F(1, 242) = 0.39$, $p = .53$ and $F(1, 242) = 0.01$, $p = .91$, for Trials 4 & 5, respectively]. There were no other significant effects [Trial: Wilks' Lambda = 0.99, $F(2, 241) = 0.675$, $p = .51$; Trial X Basic Knowledge: Wilks' Lambda = 0.98, $F(2, 241) = 1.96$, $p = .14$]. Thus, basic knowledge significantly predicted performance in all task trials. However, the test used to measure basic knowledge predicted early task performance differentially for Blacks and Whites and therefore, should be considered unfair.

Race x Understanding interaction effects on task performance. Finally, I examined the fairness of the understanding test in predicting task performance. Results indicated significant between subject effects for understanding [$F(1, 242) = 16.36$, $p < .0001$] but not for race [$F(1, 242) = 0.30$, $p = .58$] or Race X Understanding [$F(1, 242) = 0.50$, $p = .48$]. Thus, understanding significantly affected task performance.

Additionally, results indicated significant within subject effects for Trial X Race [Wilks' Lambda = 0.968, $F(2, 241) = 3.90$, $p < .05$]. Follow-up analyses indicated that Blacks performed lower than Whites in all three task trials, and that difference was larger in Trial 4 [Blacks: $M = 362.04$, $SD = 662.23$, $M = 647.22$, $SD = 844.30$, $M = 879.20$, $SD = 885.19$; Whites: $M = 935.63$, $SD = 862.53$, $M = 1342.62$, $SD = 814.87$, $M = 1540.93$, $SD = 773.90$, for Trials 3, 4 and 5, respectively]. There were no other significant effects [Trial: Wilks' Lambda = 0.98, $F(2, 241) = 2.57$, $p = .08$; Trial X Understanding: Wilks' Lambda = 0.99, $F(2, 241) = 1.25$, $p = .28$; Trial X Race X Understanding [Wilks' Lambda = 0.98, $F(2, 241) = 2.54$, $p = .08$]. Thus, understanding significantly predicted performance in all task trials and the test used to measure understanding showed no signs of differential prediction and therefore should be considered fair.

Study 2

As stated earlier, in Study 1, I did not implement a direct measure of stereotype threat. I omitted this direct measure intentionally because of the possibility that including a direct measure could make stereotype threat salient to the participants and that any effects of stereotype threat could then carry over to influence other outcomes of the study. However, theoretically, stereotype threat effects may cause the test score differences between Blacks and Whites.

Therefore, I designed a second study to assess directly the incidence of perceived stereotype threat relative to cognitive ability and the job knowledge tests used in Study 1. That is, I designed the second study to measure stereotype directly relative to both types of written tests. Furthermore, this study included the assumption that the test of understanding is the better of the two job knowledge measures. Thus, I included only the understanding test for comparison to the cognitive ability test for its induction of stereotype threat perceptions.

Furthermore, as stated earlier also, and consistent with previous research, cognitive ability tests generally produce significant black-white mean score differences. If these differences are, in fact, the result of perceived stereotype threat, then whereas Blacks should perceive the tests as threatening, whites should not have the same perceptions. This thinking led to my first hypothesis.

H_1: Blacks will perceive the tests as more (stereotype) threatening than Whites.

Further, as stated earlier also, because it is cognitive ability tests that generally produce significant black-white mean score differences, if these differences are the result of perceived stereotype threat then the cognitive ability test itself is likely causing those perceptions. However, as argued earlier, a job knowledge test should not be perceived as threatening and should therefore produce weaker stereotype threat perceptions.

H_2: The cognitive ability test will produce stronger perceptions of stereotype threat relative to the understanding test.

Method

Participants

Participants (N = 102) were undergraduate students at a medium-sized mid-western university. Of these, I obtained 74 from a pool of undergraduate students enrolled in introductory psychology classes. I recruited the remaining 28 participants from advanced psychology classes. All students participated in exchange for extra credit points. Of these, 41 were African Americans (10 males, 31 females), 57 were White Americans (21 males, 36 females), 3 were Asian Americans (females), and 1 was an Hispanic American (female).

Task Description

Participants performed an adaptation of an existing computer-based truck dispatching simulation task (Steele-Johnson & Perlow, 1989). This task required participants to receive, process, and ship orders of military parts and supplies to various areas within an Area of Responsibility (A.O.R.). Eight rules constrained the processing of these orders. Rules included information about identifying order types, order delivery schedules, truck capacity restrictions, and delivery area restrictions. Running task time was identified as the time spanning one working week. That is, the task time began at 9:00 a.m. on Monday and ended at 5:00 p.m. Friday. Each hour interval in the task was equivalent to 15 seconds in real time. Each day in the task corresponded to two minutes in real time. Each task trial lasted 10 minutes. All participants received task instructions prior to attempting the task.

Measures

Cognitive ability. I assessed cognitive ability using the first 10 questions of the *Wonderlic* Personnel Test (Wonderlic, C., 2002) used in Study 1. (See Study 1, p. 41 for sample test items and descriptive information for the test.) I administered the test in a paper-and-pencil format. I did not calculate test score because the test was used only as a reference to generate test perceptions.

Understanding. I assessed understanding using the first 10 questions of the understanding (job knowledge) test used in Study 1. (See Appendix B) I administered the test in a paper-and-pencil format. I did not calculate test score because the test was used only as a reference to generate test perceptions.

Stereotype threat. I assessed stereotype threat was assessed using an 8-item scale developed by Ployhart, Ziegert, & McFarland (2003). I administered the scale in a paper-and-pencil format. Participants rated the degree to which they agreed with each question on a 7-point Likert-type scale with 1 representing “Strongly Disagree” and 7 representing “Strongly Agree”. Ployhart et al. (2003) reported an internal reliability coefficient of .81 for this measure.

Test perceptions. I assessed test perceptions using the paper-and-pencil measure designed for Study 1. The two items used in Study 2 addressed the perception of a test as a measure of intelligence and as a measure of knowledge (See Appendix C in Study 1 for more information.). Participants rated the degree to which they agreed with each question on a 7-point Likert-type scale with 1 representing “Strongly Disagree” and 7 representing “Strongly Agree”.

Demographics. I collected demographic information after administering both of the tests, using the paper-and-pencil measure designed for Study 1 (See Appendix D). Demographic items included age, race, gender, college ranking (freshman, sophomore, junior, senior), and college major.

General perceptions. I assessed general perceptions about the fixedness (or malleability) of both intelligence and job knowledge using the paper-and-pencil measure designed for Study 1. (See Study 1, p. 45 for more information.)

Test Order Manipulation and Procedure

To examine the possibility that exposure to one test triggers stereotype threat and that this threat influences performance on the next test, I implemented two testing conditions. More specifically, it may be the case that individuals perceive the first test as threatening and then, once activated, this threat perception carries over to influence performance on the second. Therefore, I divided participants into two groups such that one half received the cognitive ability test, followed by the understanding test; the second group received the tests in the opposite order. Figure 2 shows the order of administration of experimental procedures for both testing conditions.

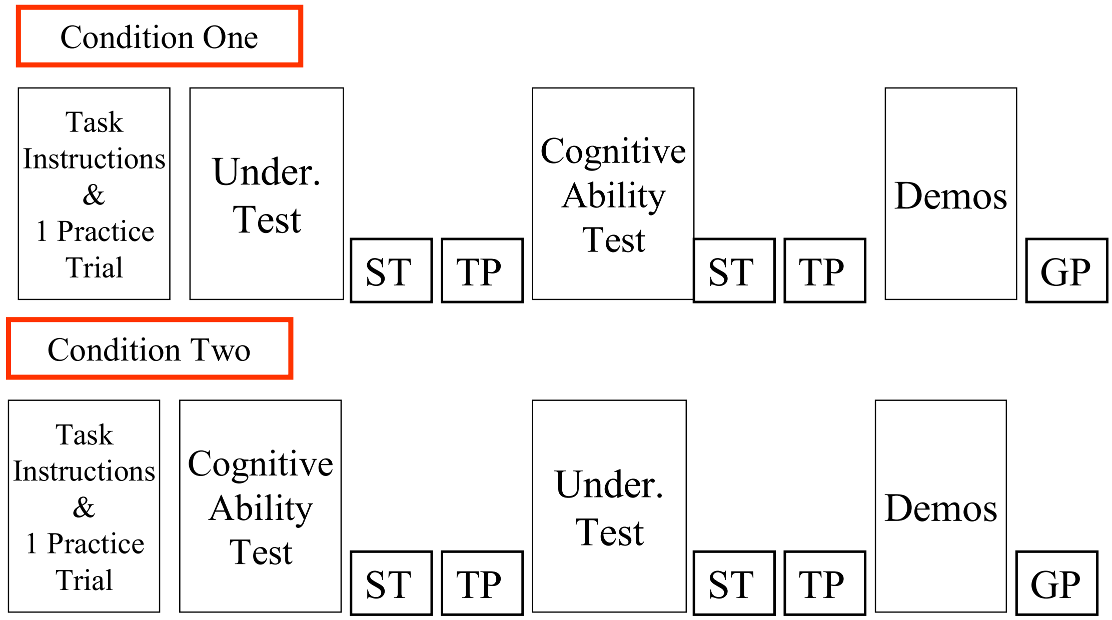

Figure 2: Experimental Procedure for Study Two

To ensure that participants could perceive the relevance (or lack thereof) of the two tests (understanding and cognitive ability) to the task, each participant received task instructions and performed one 10-minute trial of the task. Next, participants completed the written measures of cognitive ability and task understanding, with the order depending on condition. Immediately after completing the first test, participants completed the measures of stereotype threat (ST) and of test perceptions (TP) for that test. Then, participants completed the second test followed immediately by another measure of stereotype threat (ST) and test perceptions (TP). Finally, participants completed the demographics questionnaire and the general perceptions (GP) measure at the end of the experimental session. Note: Providing demographic information only at the end of the session eliminated the possibility of making race salient to the participants prior to their completion of knowledge or ability measures.

Results

Sample Characteristics

Participants (N = 102) were run in one of two experimental conditions which differed only in the order of administration of the two (cognitive ability and job knowledge) tests. In Condition 1, the understanding test preceded the cognitive ability test. In Condition 2 the test order was reversed. I observed no missing values for any test or survey response options. Further, I examined the data for the presence of possible outliers (responses beyond the possible scale) and/or response bias but observed none. Therefore, all individual data were available for use in the analyses.

In order to compare even numbers of African American participants to Whites, I had to reduce the sample. I accomplished this reduction by eliminating all 3 Asian American participants and the 1 Hispanic American participant. Then I used a random number generator to select participants to remove on the basis of race. Of the resulting sample (n = 80), 40 were African American, 40 were White with 20 from each subgroup per testing condition. I used this sample in all analyses comparing Blacks to Whites.

Means, Standard Deviations, and Intercorrelations between Study Variables

Means, standard deviations, and intercorrelations between Study 2 variables are shown in Table 4.

Table 4
Means, Standard Deviations, and Correlations of Study 2 Variables

Variable	Mean	SD	1	2	3	4	5	6	7	8	9
1 Race	-----	-----	--								
2 TPCAT1	4.87	1.53	.02	--							
3 TPCAT2	3.25	1.96	-.19	-.13	--						
4 TPUND1	3.79	1.94	-.17	.25	.07	--					
5 TPUND2	5.23	1.71	.07	.29*	-.15	.27*	--				
6 STPCAT	2.99	0.94	-.43†	.01	.18	.19	-.07	--			
7 STPUND	2.98	0.94	-.51†	-.04	.21	.29**	-.01	.85†	--		
8 GenPerc1	2.69	1.79	-.09	-.02	.03	.19	-.17	-.02	.13	--	
9 GenPerc2	6.35	1.02	-.09	.20	-.13	-.16	.15	.06	-.10	-.13	--

* $p < .05$, ** $p < .01$, *** $p < .001$, † $p < .0001$.

Key:

TPCAT1	Perceptions of the Cognitive Ability Test as a measure of intelligence
TPCAT2	Perceptions of the Cognitive Ability Test as a measure of knowledge
TPUND1	Perceptions of the Understanding Test as a measure of intelligence
TPUND2	Perceptions of the Understanding Test as a measure of knowledge
STPCAT	Perceptions of Stereotype Threat Associated with the Cognitive Ability Test
STPUND	Perceptions of Stereotype Threat Associated with the Undrsatnding Test
GenPerc1	Perceptions of Intelligence as fixed, unchangable
GenPerc2	Perceptions of Job Knowledge as Malleable

Initial Group Differences

To check for initial group differences, I compared the means for the variables within each condition (i.e., Test Order) subgroup. As stated earlier, in both groups, 50% of the participants were African American and 50% were White. Also, in Condition 1, 27 participants were female, (11 White, 16 Black) 13 were male (9 White, 4 Black) compared to 28 females (13 White, 15 Black) and 12 males (7 White, 5 Black) in Condition 2. Further, the age ranges for the two groups were similar (i, e., participants in Test Order 1 ranged in age from 17 to 24, $M = 19.5$ years; and from 17 to 32, $M = 20.05$ years in Test Order 2). Therefore, I concluded that minimal differences existed between the two groups and that race, age and gender were evenly distributed across both test order conditions.

Analysis Approach

To examine my hypotheses and in exploratory analyses, I performed a series of repeated measures anovas with condition (test order), race, and test type (ability, knowledge) as predictors and perceptions as outcomes. Test type was the only within subject factor. I included condition in the analysis to examine and control for the effects of test order on perceptions.

Effects of Condition, Race, and Test Type on Stereotype Threat Perceptions (Hypotheses 1 and 2)

Results indicated a significant main effect of Race [$F(1, 76) = 24.65$, $p < .0001$] such that Blacks [$M = 3.42$, $SD = 0.91$ perceived the tests as more threatening than Whites [$M = 2.55$, $SD = 0.66$]. Thus, Hypothesis 1, which stated that "Blacks will perceive the tests as more (stereotype) threatening than Whites." was supported. However, results indicated no significant effect for Test Type on stereotype threat perceptions [Wilks' Lambda = 0.999, $F(1, 76) = 0.472$, $p = 0.829$]. The tests were perceived as similarly threatening [$M = 2.99$, $SD = 0.94$ for stereotype perceptions for the cognitive ability test, and $M = 2.98$, $SD = 0.94$, for stereotype perceptions associated with the understanding test]. Therefore, Hypothesis 2, which stated that "The cognitive ability test will produce stronger perceptions of stereotype threat relative to the understanding test." was not supported. There were no other significant effects (see Table 5).

Table 5
Effects of condition, race, and test type on stereotype threat perceptions

Predictor	Wilks' λ	F(1, 76)	p
Between Subjects:			
Condition		3.40	0.07
Race		24.65†	--
Condition X Race		0.07	0.79
Within Subjects:			
Test Type	0.99	0.47	0.83
Test Type X Race	0.98	1.56	0.22
Test Type X Condition	0.98	1.56	0.22
Test Type X Race X Condition	0.99	0.75	0.39

† $p < .0001$

Effects of Condition, Race, and Test Type on Test Perceptions

To better understand the influence of race and test type on perceptions relevant to stereotype threat, I examined the effects of these variables on perceptions of the tests as measures of intelligence and as measures of knowledge, using the analysis approach described above and again controlling for the effects of test order. I then repeated these analyses to examine effects on perceptions of intelligence as fixed and of knowledge as malleable.

Effects of condition, race, and test type on perceptions of the tests as measures of intelligence. In examining effects on the perception of the tests as a measure of intelligence, results indicated a significant effect for test type [Wilks' Lambda = 0.79, $F(1, 76) = 20.41$, $p < .0001$, $M = 4.88$, $SD = 1.53$ and $M = 3.78$, $SD = 1.94$, for the cognitive ability test and the understanding test, respectively]. There were no other significant effects (see Table 6).

Table 6
Effects of condition, race, and test type on perceptions of the tests as measures of intelligence.

Predictor	Wilks' λ	*F*(1, 76)	*p*
Between Subjects:			
Condition		2.53	0.11
Race		1.04	0.31
Condition X Race		0.37	0.54
Within Subjects:			
Test Type	0.788	20.41†	--
Test Type X Race	0.971	2.27	0.14
Test Type X Condition	0.998	0.13	0.72
Test Type X Race X Condition	0.997	0.22	0.64

† $p < .0001$

Effects of condition, race, and test type on perceptions of the tests as measures of knowledge. In examining effects on perceptions of the tests as measures of knowledge, results indicated a significant effect for test type [Wilks' Lambda = 0.66, $F(1, 76) = 39.88$, $p < .0001$, $M = 3.25$, $SD = 1.96$ and $M = 5.22$, $SD = 1.71$, for the cognitive ability test and the understanding test, respectively]. Thus, participants perceived the understanding test as a stronger measure of knowledge than the cognitive ability test. Also, results indicated a significant main effect for Condition [$F(1, 76) = 11.35$, $p < .01$]. There were no other significant effects (see Table 7).

Table 7
Effects of condition, race, and test type on perceptions of the tests as measures of knowledge.

Predictor	Wilks' λ	$F(1, 76)$	p
Between Subjects:			
Condition		11.35**	--
Race		0.98	0.32
Condition X Race		0.80	0.38
Within Subjects:			
Test Type	0.66	39.88,†	--
Test Type X Race	0.97	2.56	0.11
Test Type X Condition	0.99	0.10	0.75
Test Type X Race X Condition	0.99	0.06	0.81

$^{**}p < .01$, † $p < .0001$

To explore the effect of condition, I examined the means for the two tests within each condition. Results indicated that, regardless of test type, providing the understanding test first (Condition 1) resulted in perceiving the tests more strongly as measures of knowledge [$M = 4.66$, $SD = 1.06$] and providing the cognitive ability test first (Condition 2) resulted in perceiving the tests less strongly as measures of knowledge [$M = 3.81$, $SD = 1.19$].

Effects of condition and race on perceptions of intelligence as fixed and/or knowledge as malleable. First, I examined the effects of condition and race on the perception of intelligence as fixed (see Table 8). Results indicated no significant effects of Condition [$F(1, 76) = 0.32$, $p = 0.57$], Race [$F(1, 76) = 0.67$, $p = 0.42$], or Condition X Race [$F(1, 76) = 2.87$, $p = 0.09$]. Next, I examined the effects of condition and race on the perception of knowledge as malleable (see Table 8). Again, there were no significant effects of Condition [$F(1, 76) = 0.00$, $p = 1.00$], Race [$F(1, 76) = 0.76$, $p = 0.38$], or

Condition X Race [$F(1, 76) = 1.72$, $p = 0.19$]. Thus, all participants failed to report the perception that intelligence was fixed and all participants viewed their knowledge as malleable.

Table 8
Effects of condition and race on perceptions of:
Intelligence as Fixed and/or Job Knowledge as Malleable

Predictor	*F*(1, 76)	*p*	Cond1 *M*, *SD*	Cond2 *M*, *SD*	Blacks *M*, *SD*	Whites *M*, *SD*
Intelligence-Fixed:						
Condition	0.32	0.57	2.80, 1.94	2.58, 1.64		
Race	0.67	0.42			2.85, 2.01	2.52, 1.55
Condition X Race	2.87	0.09				
Knowledge-Malleable:						
Condition	0.00	1.00	6.35, 2.58	6.35, 0.80		
Race	0.76	0.38			6.45, 0.81	6.25, 1.19
Condition X Race	1.72	0.19				

DISCUSSION

Study 1 Discussion

Purpose

The overall purpose of Study 1 was to examine the possibility of reducing adverse impact in selection by implementing the use of job knowledge tests for predicting job performance. I accomplished this by comparing the effectiveness of two job knowledge tests for predicting task performance to that of a cognitive ability test. Also, I assessed job knowledge as a mechanism through which cognitive ability affects task performance. Further, I designed the two job knowledge tests to explore different conceptualizations of job knowledge and compared the efficacy of each for predicting performance. Finally, I examined the effects of perceived stereotype threat as a possible cause of Black-White score differences relative to each of the three (cognitive ability and both job knowledge) tests.

Contributions

This study contributed to current theory and research by integrating concepts from cognitive psychology, industrial-organizational psychology, and social psychology in an effort to improve existing selection concepts and procedures. Specifically, first this study contributed to existing selection theory and research by extending our current conceptualization of job knowledge. This extension provided an alternative method of defining and measuring job knowledge.

Second, this study extended our current understanding of the efficacy of cognitive ability and job knowledge as predictors of performance. Results indicated that both job knowledge tests predicted performance better than the test of cognitive ability. This comparison extended our current understanding of the effectiveness of using these two test types for selection.

Third, this study extended our current understanding of job knowledge as a mechanism through which cognitive ability affects performance. Results indicated that job knowledge completely mediated the ability – performance relationship. Thus, this study extended our current knowledge of the mediating effects of job knowledge in the cognitive ability – performance relationship. Therefore, it may be possible to replace cognitive ability tests with tests of job knowledge in the selection process.

Fourth, this study extended our current knowledge of proposed causes of stereotype threat perceptions. Results indicated that individuals were able to distinguish between what each test was designed to measure. Specifically, participants perceived the cognitive ability test as a measure of intelligence and the job knowledge tests as measures of knowledge. Further, results indicated that individuals failed to perceive their intelligence as a fixed, stable characteristic. Thus, this study extended current theoretical conceptualizations relative to the proposed cause of perceived stereotype threat.

Finally, this study extended our current knowledge of perceived stereotype threat effects on differences in Black-White test score performance. Results indicated no significant reduction of Black-White test score differences for either job knowledge test relative to the cognitive ability test. Therefore, this study extended existing theory relative to selection tests by comparing the Black-White score differences on each type of test.

Theoretical Issues and Implications

One theoretical issue that must be addressed is the extension of our current theoretical conceptualization of job knowledge. Much of current, existing theory has identified job knowledge as including only declarative and procedural information. Thus, many researchers have focused on this concept for defining and measuring job knowledge in their research (e.g., Anderson, Greeno, Kline, & Neves, 1981; Hunter, 1993; Schmidt, Hunter, & Outerbridge, 1986; Martocchio, 1994). However, others have attempted to extend this conceptualization of job knowledge, and thus, its definition and measurement techniques to include other components of knowledge (e.g., Greeno, 1989; Greeno, Riley, & Gelman, 1984; Hedlund, Forsythe, Horvath, Williams, Snook, & Sternberg, 2003; Schmidt & Hunter, 1993; Sternberg, 1985; Greeno & Simon, 1988; and Bell & Kozlowski, 2002a, 2002b).

The results of the current study have demonstrated the importance of the extension of the concept of job knowledge within the theoretical literature. In the current study, both measures of job knowledge effectively predicted performance. However, when job knowledge was viewed as declarative and procedural information only, "basic knowledge," the test that was developed using this concept exhibited differential prediction for the two racial subgroups. Alternatively, the concept of job knowledge as including other knowledge components led to the development of a test that accounted for *slightly* more variance in performance than the basic knowledge test and demonstrated no such differential prediction. Thus, selection theory should include this extended concept of job knowledge to improve selection testing for prediction and to investigate further the possible reduction of adverse impact.

Another theoretical issue that has been a focus of the study is the mediation of the relationship of cognitive ability and performance by job knowledge. Much of the existing research has touted cognitive ability as the best predictor of performance on the job. To date, many theorists and researchers seem to have adhered to this type of cognitive ability as the 'be all, end all' model for predicting performance (e.g., Ree & Earles, 1991; Ree, Earles, & Teachout, 1994).

However, some research has identified job knowledge as a mechanism through which cognitive ability may affect performance. This line of research, mostly investigated by Hunter, Schmidt, and their colleagues, throughout the 1980s and 1990s, has attempted to examine job knowledge as a mediator in the cognitive ability – performance relationship. In that earlier mediation literature, however, whether due to the sample used, or the type of task, or the technique for measuring job knowledge, the authors reported only partial mediation of the relationship (e.g., Hunter, 1986, Hunter & Schmidt 1996), suggesting both direction and indirect effects.

The results of the current study extend that theoretical mediation literature by demonstrating that, regardless of how job knowledge was measured, complete mediation was the result. Thus, the current research suggests an indirect effect of cognitive ability on performance through knowledge acquisition (job knowledge), thereby extending Hunter's (1983) model. According to these results, it seems clear that ***cognitive ability affects performance through its effect on knowledge acquisition***. Thus, a job knowledge mediation model is a better theoretical model than that which focuses on cognitive ability alone as a predictor of performance.

Practical Implications

In a practical sense, (although important for theory as well) these results suggest that tests of job knowledge *may* predict performance better than tests of cognitive ability. Specifically, both job knowledge tests accounted for significantly more variance in performance than the cognitive ability test. Thus, job knowledge tests could be used to replace cognitive ability tests in selection. The result would be better prediction of future job performance.

Further, I would argue that a test of understanding may be more beneficial in selection than either a test of basic knowledge or cognitive ability. This idea is supported by the finding that a test of understanding accounts for more variance in performance than a basic knowledge test. However, it could be suggested that these results demonstrated that, because the increase in variance accounted for by the understanding test over that accounted for by the basic knowledge test was not significant ***statistically***, either type of job knowledge test would be acceptable for use in selection.

In response, I would argue that the difference of 4% of the variance accounted for was significant in a practical sense. Given the cost of selection and training and the additional costs associated with re-selection and training for replacements for employees who were incorrectly hired based on *faulty* selection procedures, it seems that utility must be considered. This idea has support in existing research. For example, when considering test fairness and utility, Hunter, Schmidt, and Rauschenberger (1977) proposed that, "Because no accepted method for converting minority selection ratios to utility units exists, … it was concluded that each personnel researcher or organization must consider the trade-off between utility and the minority selection ratio subjectively…" (p. 245). Thus, I would argue that considering the additional costs associated with 'mis-hiring' even one individual who would later either leave the job voluntarily or be released for poor performance is one practical way of considering utility for the company.

Further, I would argue that an additional benefit with the use of the understanding test relates to its ability to predict performance uniformly. Specifically, whereas the basic knowledge exhibited differential prediction for Blacks and Whites on earliest task performance, results of this study demonstrated that the understanding test exhibited no such differential prediction. Given that the purpose of testing is to identify those individuals who will acquire the relevant, necessary job knowledge earliest, and be able to perform the job soonest, the focus should be on predicting early task performance. Using a test that predicts early task performance differentially, such as the basic knowledge test, would necessarily invite the possibility of causing adverse impact. Therefore, the understanding type of test should be recommended for selection.

In sum, based on these results, it may be possible to use a test of (job) understanding in place of a cognitive ability test, in specific situations, for employee selection, placement, or for promotion. Admittedly, the use of any job knowledge test in selection requires the applicants to have some experience with the job. The results of the current study, though, demonstrated that, even with minimal exposure to the task at hand, individuals can and do acquire job knowledge and that a test of that knowledge is a good predictor of future performance. Thus one possible setting where the use of this test could be advantageous is in an assessment center. Another is in choosing incumbents for promotion, as was the case in the Ford example. Finally, this test could be beneficial in any setting where existing measures have been found to result in adverse impact.

Future Research

In study 1 I examined the differences in perceptions relative to the cognitive ability test and the two job knowledge tests for Blacks and Whites and differences in the test scores for each racial subgroup. However, other racial and non-racial subgroups are protected from discrimination under Title 7 as well. Further, previous research has identified test score differences between the majority and other protected subgroups (e.g., women) (e.g., McIntyre, Paulson, & Lord, 2003; Martens, Johns, Greenberg, & Schimel, 2006; Marx & Roman, 2002; Pinel, 1999). Therefore, along with extended research which compares other combinations of racial sub-groups for differences on these tests, future research should include an examination of test score differences for males and females on job knowledge tests as well. Also these examinations should include a comparison of the efficacy of cognitive ability and job knowledge tests for predicting performance for males vs. females.

Study Limitations

One limitation of Study 1 was that the sample included university students only. One could suggest that this sample was 'more well-educated' than the population as a whole. However, I would argue that the open enrollment policy of this university provided the opportunity to collect a sample of participants with a wide range of variability in backgrounds, experiences, and abilities. For example, in this sample, cognitive ability scores were normally distributed, ranging from a score of 7 on the lower end to a top score of 35 [*M* = 21.2, *SD* = 5.8]. This range is similar to that observed for employees in general [*M* = 21.7, *SD* = 7.6] (Wonderlic, 2002). Furthermore, this student sample was not limited to introductory psychology students or to undergraduate students as is often the case in university research. Rather, this sample included students from a wide variety of university majors, from a wide variety of advanced courses, and from all levels of college experience (freshman through graduate students) as well. So, I would argue that the current sample was more representative of the University population and because of the open enrollment policy more representative of the general population than is often the case in studies using a college student sample. I would argue further that because of the variability in backgrounds and experiences, the current sample was more representative of the general population, that this sample was more likely to represent those that enter the workforce and, therefore, that these results should be more generalizable to that workforce population.

A second limitation of Study 1 was that it focused on predicting performance on a single, non-changing task only. That is, the static task used in this study provided no opportunity for the measurement of, or demonstration of, decision making performance in a dynamic job environment. Thus, the tests of "job" knowledge used in this study were, in actuality, tests of "task" (specific) knowledge and therefore were limited to sampling knowledge of and predicting performance relative to this one static task alone. It may be the case that, in a dynamic task environment, where the performance criteria are changing, a cognitive ability test would be better able to predict performance than a test of specific task-related knowledge.

Finally, a third limitation of Study 1 was that the study did not include a direct measure of stereotype threat perceptions relative to the written tests. This was an intentional omission because, according to theory (e.g., Steele & Aronson, 1995), simply making race a salient factor in the study could trigger stereotype threat and thereby alter the results. A direct measure of perceived stereotype threat would force race to become salient to all participants. Therefore, I implemented Study 2 to assess stereotype threat perceptions directly.

Study 2 Discussion

Purpose

The first purpose of Study 2 was to examine directly the occurrence of perceived stereotype threat relative to a cognitive ability and a job knowledge test. A second purpose was to examine two of the proposed theoretical causes of this 'threatening' perception. One objective was to extend our current knowledge of stereotype treat effects by testing the theoretical assertion that stereotype threat may cause black-white test score differences. I accomplished this by examining perceived stereotype threat relative to a cognitive ability test and a job knowledge (understanding) test. The second objective was to extend stereotype theory by testing the proposal that Blacks experience stereotype threat because of perceptions relative to the tests themselves. I accomplished this by testing 1) whether participants recognized that the cognitive ability test was designed as a measure of intelligence and that the understanding test was designed as a measure of job knowledge, and 2) whether participants perceived that their intelligence was a fixed characteristic. Pursuing these objectives allowed us to extend our current understanding of the stereotype threat construct.

Contributions

Study 2 was designed to contribute to current theory and research by examining, first, whether Blacks held stronger perceptions of stereotype threat than Whites relative to the two tests. Results indicated that Blacks perceived *both* tests as more threatening than Whites. Thus, this study contributed to theory by extending our current knowledge regarding the types of test on which stereotype threat may have an impact.

Second, this study was designed to contribute to current theory by examining the possibility that perceiving a test as measure of intelligence was a possible cause of stereotype threat. Results indicated that Blacks and Whites perceived the cognitive ability test as a stronger measure of intelligence than the understanding test. Therefore, this study contributed to existing theory by extending our current knowledge of the relationship between the perception of the test as a measure of *intelligence* and the perception of stereotype threat.

Third, this study extended our knowledge about the perception that a test is a measure of (job) knowledge. Results indicated that Blacks and Whites perceived the understanding test as a stronger measure of job knowledge than the cognitive ability test. Therefore, this study contributed to existing theory by extending our current knowledge of the relationship between the perception of the test as a measure of *knowledge* and the perception of stereotype threat.

Finally, Study 2 extended our current understanding of theory by examining whether Blacks perceived their intelligence as a fixed, stable, individual characteristic. Results indicated the Blacks and Whites failed to report the perception that intelligence is fixed. Thus, this study contributed to theory by extending our current knowledge about the relationship between the perception of intelligence as fixed and the perception of stereotype threat.

Theoretical Issues and Implications

To reiterate, Steele and Aronson (1995) proposed stereotype threat as a possible cause of the Black-White score differences on cognitive ability tests. Stereotype threat theory suggests that test score differences are the outcome of knowledge, perceptions, and beliefs. Specifically, according to the theory, Blacks hold previous knowledge of the stereotype that members of their racial subgroup generally performed lower than Whites on intelligence tests. Also, the theory proposes that Blacks perceive cognitive ability tests as measures of intelligence and that they believe also that their intelligence is a fixed trait. Further, according to the theory, this combination of perception and belief causes Blacks to feel the threat that they would confirm that racial stereotype by exhibiting poorer test performance than Whites. Finally, it is the experience of this threat that results in lower test scores for Blacks on cognitive ability tests.

However, Steele and Aronson's (1995) theory did not address the possibility that stereotype threat might cause Black-Whites score differences on job knowledge tests. Therefore, I tested whether Black-White differences in perceptions existed for the two different types of tests. First, I tested whether perceptions of what the test measured differed for Blacks and Whites for a cognitive ability test. Second, I tested whether the same perceptions differed between the two groups for a job knowledge test. Finally, I tested whether stereotype threat was perceived by Black and Whites for each of the two tests.

Results indicated that Blacks perceived the tests as more threatening than Whites. However, contrary to my hypothesis, this threat was perceived for both the job knowledge *and* the cognitive ability test. Further, all participants perceived the cognitive ability test as a stronger measure of intelligence than the job knowledge test and the understanding test as a stronger measure of job knowledge. Also, and possibly most importantly, administering the understanding test before the cognitive ability test "altered" perceptions associated with the cognitive ability test such that the latter was perceived more strongly as a measure of job-related knowledge *rather than intelligence.* Finally, all participants failed to report the view that intelligence was a fixed, stable trait and all reported the view that job knowledge was malleable. Thus, the theoretical proposal that the combination of perceiving the test as measuring intelligence and viewing intelligence as fixed produced stereotype threat (Steele & Aronson, 1995; Steele, 1997) failed to be supported by these results.

Thus, the theoretical implications of the results of Study 2 are that there may be limitations to existing stereotype theory. That is, the theoretically proposed causes of stereotype threat perceptions have not been supported in this study. Specifically, I found that both Blacks ***and*** Whites perceived the cognitive ability test as a measure of intelligence. Thus, these results have demonstrated that this type of perception was not limited to Blacks and therefore, this study failed to support the assertion that perception of the test as a measure of intelligence is a possible trigger of stereotype threat for Blacks.

Additionally, I found that both Blacks ***and*** Whites perceived the understanding test as a measure of knowledge, yet the test was perceived still as threatening by Blacks. Thus, this study contributed to existing theory by demonstrating that, even when the test is perceived to measure something other than of intelligence, stereotype threat can still occur. Therefore, the theory should include an assessment of the possibility that this type of perception may be related to stereotype threat also.

Finally, I found that neither Blacks nor Whites perceived their intelligence as a fixed characteristic. Thus, again, the results have failed to support existing theory by demonstrating that perceptions of intelligence as fixed should not be considered as a possible trigger of stereotype threat for Blacks. Therefore, these results have eliminated the possibility of this type of perception from the list of possible causes of stereotype threat for Blacks as well.

Taken together, these results have clarified specifics of the existing theory regarding the occurrence of stereotype threat. Further, these results have suggested that there must, necessarily, be other factors that produce stereotype threat perceptions for Blacks. Therefore, this study extends our current conceptualization regarding specifically proposed causes of stereotype threat for Blacks and the perception of stereotype threat relative to specific selection tests.

Practical Implications

The results of Study 2 provide implications for practice in the area of selection. Specifically, these results suggest that either type of test can produce perceptions of stereotype threat for Blacks. More specifically, the practice of replacing cognitive ability tests with job knowledge tests for selection will not alter Blacks' perceptions of the test as being threatening. Thus, rather than directing our focus on reducing or eliminating stereotype threat relative to testing, practitioners may want to examine other constructs that have been proposed for the reduction of adverse impact in selection. These could include "cultural identification" (Helms, 2002), non-cognitive predictors (Goldstein, Zedeck, & Goldstein, 2002), and/or composite predictors (Baron & Janmen, 1996; Sackett & Ellingson, 1997).

Future Research

In Study 2 I focused specifically on stereotype threat theory. Results demonstrated that existing theory of perceived stereotype threat has limitations. Specifically, these results suggest that the proposed causes of this threat may be in error. That is, the results of Study 2 have demonstrated that, whether we use a cognitive ability or a job knowledge test for employee selection, and no matter what the test is perceived to measure, and regardless of whether intelligence is viewed as a stable characteristic, Blacks will perceive the test as threatening. Therefore, future research should examine other possible ‘predictors’ of stereotype threat perceptions. One suggestion might be to examine threat associated with the test administration format itself. That is, it could be the case that the paper-and-pencil testing format is causing the threat. Future research could compare this format to a test administered via computer.

Study Limitations

This study was limited in that the sample included undergraduate students only. However, as stated earlier, this group was not limited to introductory psychology students, as is often the case in university research. Rather, this sample included students from advanced courses as well. Therefore, I would argue that this sample was more representative of the University and the general population than is often the case, and therefore, that these results are more generalizable to the population as a whole.

General Discussion

Taken together, these two studies have integrated insights from previous research to examine issues relative to and extend our understanding of the use of tests in selection. Job knowledge has been identified as a better predictor than cognitive ability. Therefore, a test of job knowledge could be used in place of cognitive ability tests for employee selection.

Further, the results of these studies enabled us to extend our concept of job knowledge beyond that of basic facts and procedures to include more 'abstract' decision-making information. We can design job knowledge tests that include that more abstract information as well. The end result is a test of job-related 'understanding'.

Also, these results have supported the idea that cognitive ability effects performance through job knowledge. Job knowledge completely mediated the relationship. Thus, theory should focus on this mediation model.

Finally, the results of both of these studies suggest that we should re-examine stereotype threat theory. First, (as stated above) investigators should re-examine the theory to identify actual inducers (predictors) of stereotype threat. Second, because results demonstrated no changes in the range of score differences for Blacks on the three tests, stereotype threat should be examined more closely to determine its actual effects on test performance. Finally, researchers should reinvestigate the effects of perceived stereotype threat on job performance. In general, results have indicated that Blacks perform lower than Whites. Isn't this what we should consider most important? The bottom line is: Does the perception of stereotype threat on selection tests affect final performance on the job? According to the results found in these two studies, the answer is unclear.

In sum, the results of these studies have suggested that there may exist other causes of stereotype threat than those proposed by current theory. Specifically, perceptions of what is measured by the test (e.g., intelligence or knowledge) seem to have no effect on perceived threat. The same holds true for perceptions of intelligence as a fixed or malleable construct. Stereotype threat may exist, and it may affect performance on selection tests, and perhaps even job performance. However, the theoretical background behind the theory seems to be lacking and therefore, must be reinvestigated and redefined. In fact, the results of these two studies together suggest that, until stereotype threat theory is modified and/or clarified, perhaps we should omit this theory from our quest to reduce adverse impact, rather than directing our focus on reducing or eliminating stereotype threat relative to testing.

Conclusion

In conclusion, the results of the current studies have demonstrated that stereotype threat theory must be revisited before it can be beneficial in selection. However, this was but one of the focuses of the current research. Other objectives have produced results that have demonstrated that we can replace cognitive ability tests with job knowledge tests in selection. Also, this research has shown that we can and should implement the use of a test of job *understanding*. This implementation will allow us to select employees using a better test: one with better prediction and that allows for selection of the best performers, at the outset of their appointment on the job. Further, we can use a test that results in fewer selection errors. This translates into lower costs due to incorrectly selecting the 'wrong' individuals. Finally, this research has demonstrated that, because this type of test is free from differential prediction, it is possible to use an understanding test to decrease the incidence of adverse impact in selection. Using an understanding test provides protection from A I. lawsuits because the test is absolutely job-related, and therefore, legally defensible. Therefore, using a test of job understanding in employee selection will result in $$$ savings for the organization.

Footnotes

[1]The concept of general cognitive ability or psychometric *g* was first proposed by Galton in the late 1800s, and Spearman is credited with early analytical use of psychometric *g* (in Ree & Earles, 1991).

[2]To clarify, this is not intended as a statement about the effectiveness of Cardio-Pulmonary Resuscitation, rather, an explanation of a 'proceduralized' task. I understand that the success of C.P.R. depends on several factors including, but not limited to the time between the onset of the attack and discovery of the victim, access to medical attention, and the overall health of the victim.

REFERENCES

Anderson, J. R., Greeno, J. G., Kline, P. J., and Neves, D. M. (1981). Acquisition of problem-solving skill. In Anderson, J. R. (Ed.), *Cognitive Skills and Their Acquisition* (pp. 191-230). Hillsdale, NJ: Lawrence Erlbaum Associates.

Aronson, J., Fried, C. B., and 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..

Barrett, G. V., Carobine, R. G., and Doverspike, D. (1999). The reduction of adverse impact in an employment setting using a short-term memory test. *Journal of Business and Psychology*, *14*(2), 373-377.

Baron, R. M., and Kenny, A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Bell, B. S. and Kozlowski, S. W. J. (2002a). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology, 87,* 497-505.

Bell, B. S., and Kozlowski, S. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, *55*(2), 267-306.

Brown, R. P., and Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*(6), 626-633.

Cascio, W. F. (1998). *Applied Psychology in Human Resource Management, 5th ed.*. Upper Saddle River, N.J.: Prentice Hall.

Cascio, W. F., and Phillips, F. (1979). Performance testing: A rose among thorns? *Personnel Psychology*, *32*(4), 751-766.

Chan, D., Schmitt, N., Sacco, J. M., and DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology, 83(3),* 471 – 485.

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., and Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology, 82(2),* 300 – 310.

Chi, M., Feltovich, P. J., & Glaser, R. (1996). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Chung-Yan, G. A., and Cronshaw, F. (2002). A critical re-examination and analysis of cognitive ability tests using the Thorndike model of fairness. *Journal of Occupational and Organizational Psychology*, *75*(4), 489-509.

Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Education al Measurement, 5,* 115 – 124.

Cohen, J., and Cohen, P. (1975). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Conway, A. R. A., Kane, M. J., and Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547-552.

Dubois, D. A. and Shalin, V. L. (1995).Adapting cognitive methods to real-world objectives: An application to job knowledge testing. In P. D. Nichols, S. F. Chipman, R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 189-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dubois,D. A., Shalin, V. L., Levi, K. R., & Borman, W. C. (1997). A cognitively-oriented approach to task analysis. *Training Research Journal, 3,* 103-141.

Ellis, A. P. J., and Ryan, M. (2003). Race and Cognitive-Ability Test Performance: The Mediating Effects of Test Preparation, Test-Taking Strategy Use and Self-Efficacy. *Journal of Applied Social Psychology*, *33*(12), 2607-2629.

Farr, J. L. (2003). Introduction to the special issue: Stereotype threat effects in employment settings. *Human Performance*, *16*(3), 179-180.

Fisher, S. L., and Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learner outcomes. *Personnel Psychology, 51,* 397-420.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., and Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83,* 218-233.

Goldstein, H. W., Zedeck, S., and Goldstein, I. L. (2002). g: Is this your final answer? *Human Performance*, *15*(1-2), 123-142.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, *15*(1-2), 25-46.

Greeno, J. (1989). A perspective on thinking. *American Psychologist, 44,* 134-141.

Greeno, J. G., Riley, M. S., and Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, *16*(1), 94-143.

Greeno, J. G. and Simon**,** H. A. (1988) Problem solving and reasoning. *Stevens' Handbook of Experimental Psychology, Vol. 2: Learning and Cognition (2nd ed.),* 589-672. John Wiley & Sons, Oxford, England.

Guion, R. M. (2000). Assessment, Measurement, and Prediction for Personnel Decisions.

Hausdorf, P. A., LeBlanc, M. M., and Chawla, A. (2002). Cognitive ability testing and employment selection: Does test content relate to adverse impact? *Applied H.R.M. Research*, *7*(1-2), 41-48.

Hedlund, J., Forsythe, G. B., Horvath, J. A., Williams, W. M., Snook, S., and Sternberg, R. J. (2003). Identifying and assessing tacit knowledge: Understanding the practical intelligence of military leaders. *Leadership Quarterly*, *14*(2), 117-140.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, *47*(9), 1083-1101.

Helms, J. E. (2002). A remedy for the black-white test score disparity. *American Psychoogist.* 303-305.

Higuera, Aramburu-Zabala L. (2001). Adverse impact in personnel selection: The legal framework and test bias. *European Psychologist*, *6*(2), 103-111.

Hough, L. M., and Oswald, L. (2000). Personnel selection: Looking toward the future--Remembering the past. *Annual Review of Psychology*, *51*, 631-664.

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

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior, 29(3),* 340-362.

Hunter, J. E. (1993). A causal model of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. J. Landy, S. Zedeck, & J. Cleveland (Eds*.), Performance Measurement and Theory* (pp. 257-266). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hunter, J. E., and Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96(1),* 72-98.

Hunter, J. E., and Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, & Law, 2(3-4),* 447-472.

Hunter, J. E., Schmidt, F. L., and Rauschenberger, J. M. (1977), Fairness of psychological tests: Implications of four definitions for selection utility and minority hiring. *Journal of Applied Psychology. 62(3),* 245 – 260.

Jensen, A. R. (1995). Psychological research on race differences. American Psychologist, 50(1), 41 – 42.

Landy, F. and Vasey, J. (1991). Job analysis: The composition of SME samples. *Personnel Psychology, 44, 27-50.*

Lave, J., and Wenger, E. (1991). *Situated learning : legitimate peripheral participation.* Cambridge [England] ; New York : Cambridge University Press.

Lievens, F., and Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, *10*(4), 245-257.

Martens, A., Johns, M., Greenberg, J., and Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, *42*(2), 236-243.

Martocchio, J. J. (1994). Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. *Journal of Applied Psychology*, *79*(6), 819-825.

Marx, D. M., and Roman, S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, *28*(9), 1183-1193.

McIntyre, R. B., Paulson, R. M., and Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, *39*(1), 83-90.

McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance*, *16*(3), 181-205.

Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg, and Urbina. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51(2),* 77 – 101.

Nyborg, H.,and Jensen, R. (2000). Black-white differences on various psychometric tests: Spearman's hypothesis tested on American armed services veterans. *Personality and Individual Differences*, *28*(3), 593-599.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, *26*(3), 291-310.

Palumbo, M. V. (2004). The mediating effects of task learning in the goal orientation - task performance relationship. *Unpublished Master's Thesis.* Wright State University.

Palumbo, M. V., Miller, C. E., Shalin, V. L., and Steele-Johnson, D. (2005). The impact of job knowledge in the cognitive ability - performance relationship. *Applied Human Resource Management, 10(1),* 13-20.

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology*, *76*(1), 114-128.

Ployhart, R.E., Ziegert, J. C., and McFarland, L. A (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance 16(3),* 231 – 259.

Ree, M. J., and Earles, A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*(2), 321-332.

Ree, M. J., Earles, J. A., and Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, *79*(4), 518-524.

Rushton, J. P.,.and Jensen, R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *11*(2), 235-294.

Sackett, P. R., and Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*, 707-721.

Schmader, T., and Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*(3), 440-452.

Schmidt, F. L., and Hunter, J. E. (1982). Two pitfalls in assessing fairness of selection tests using the regression model. *Personnel Psychology, 35,* 601 – 607.

Schmidt, F. L., and Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, *2*(1), 8-9.

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

Schmidt, F. L., Hunter, J. E., and Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, *71*(3), 432-439.

Smith, D. A., Greeno, J. G., and Vitolo, T. M. (1989). A model of competence for counting. *Cognitive Science: A Multidisciplinary Journal*, *13*(2), 183-211.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613-629.

Steele-Johnson, D. and Perlow, R. (1989). Computer-generated Truck Dispatch Simulation Task. University of Houston.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence.* New York, N. Y.: Cambridge University Press.

Tenopyr, M. L. (2002). Theory versus reality: Evaluation of g in the workplace. *Human Performance*, *15*(1-2), 107-122.

Civil Rights Act of 1964. Tower Amendment, Section 703h.

U. S. Supreme Court, (1971). *Griggs v. Duke Power Company. United States Reports* 401, 402.

United States District Court, Southern District Of Ohio, Western Division. (2004). *Robinson v. Ford Motor Company, Inc*. Civil Action No. 1:04-cv-00844-SAS Doc. 28.

Urbina, S. (1999). The past revisited: Black and White thinking about race and tests. *PsycCRITIQUES*, 44(1), 51-52.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*, 513-539.

Verive, J. M., and McDaniel, A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. *Intelligence*, *23*(1), 15-32.

Vernon, P. E. (1979). *Intelligence: Heredity and Environment.* San Francisco, CA: W. H. Freeman and Company.

Wernimont, P. F. and Campbell, P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*(5), 372-376.

Wonderlic, C. (2002). Wonderlic Personnel Test & Scholastic Level Exam. *Libertyville, IL: Wonderlic Personnel Test, Inc.*

Appendix A

MVPBas Learning Test A

Participant Number: __________ Date: ________________

Directions: Please circle the letter that corresponds to the answer that **best** completes the question.

IMPORTANT!!! Please DO NOT SKIP ANY QUESTIONS or leave any questions unanswered.

1. How is a rush order designated?
    a. By a "P" in the M column.
    b. By a "0" in the M column.
    c. By a "1" in the M column.
    d. By a "2" in the M column.
    e. None of the above.

2. What is accomplished by pressing CTRL-Y?
    a. You can ship a truck.
    b. You can ship the pickup order(s).
    c. You can assign orders to a specific truck.
    d. You can send pickup orders to the pickup window.
    e. None of the above.

3. A penalty is assessed if you:
    a. Assign orders to more than three trucks.
    b. Select orders for less than three zones.
    c. Ship orders to more than three zones.
    d. Accept more than 20 orders.
    e. None of the above.

4. Pickup orders must be shipped:
    a. On a truck.
    b. Either on a truck or to the pickup window.
    c. Pickup orders may not be shipped.
    d. Via the pickup window.
    e. None of the above.

5. Which of the following is true in this task?
    a. If you ship a truck that is filled with only regular orders, a penalty is assessed.
    b. If you ship a truck that exceeds minimum zone capacity, a penalty is assessed.
    c. If you ship a truck that exceeds maximum truck capacity, a penalty is assessed.
    d. If you ship a truck that is filled with both regular and rush orders, a penalty is assessed.
    e. None of the above.

6. What is accomplished by pressing the number keys?
    a. You can move between top and bottom windows.
    b. You can view the rule associated with that number.
    c. You can assign orders to trucks or the pickup window.
    d. You can see how many orders you have for processing.
    e. None of the above.

7. One day is:
    a. The time limit placed on shipping all accepted orders.
    b. The time limit placed on shipping regular orders.
    c. The time limit placed on shipping incoming orders.
    d. The time limit placed on shipping pickup orders.
    e. None of the above.

8. What key or keys allow you to ship a pickup order?
    a. CTRL-Y.
    b. CTRL-A.
    c. CTRL-X.
    d. CTRL S.
    e. None of the above.

9. What is designated by a "P" in the processing (bottom, right-hand) window.
    a. The location for assignment of truck capacity.
    b. The location for assignment of orders to a specific truck.
    c. The location for assignment of incoming orders.
    d. The location for assignment of accepted orders.
    e. None of the above.

10. What key or keys allow you to move between the top and bottom windows?
    a. The ENTER key.
    b. The F1 key.
    c. The CTRL and A keys.
    d. The CTRL and Y keys.
    e. None of the above.

11. Which of the following is true in this task?
    a. If you take too much time in reading the performance screen, a penalty is assessed.
    b. If you move between the left and right hand windows, a penalty is assessed.
    c. If you move between top and bottom windows too many times, a penalty is assessed.
    d. If you are too slow in accepting new orders, a penalty is assessed.
    e. None of the above.

12. How is a pickup order designated?
    a. By a "Y" in the M column.
    b. By a "0" in the M column.
    c. By a "1" in the M column.
    d. By a "2" in the M column.
    e. None of the above.

13. A penalty is assessed if you:
    a. Ship a whole order to fewer than three zones.
    b. Ship a whole order, either by truck or via the pickup window.
    c. Ship a pickup order to fewer than three zones.
    d. Ship a partial order, either by truck or via the pickup window.
    e. None of the above.

14. Which of the following is true in this task?
    a. Truck loads exceeding truck capacity by more than 20% may not be shipped.
    b. Truck loads must be assigned to 20% of the available delivery zones.
    c. Trucks loaded to less than 20% of maximum capacity may not be shipped.
    d. Trucks must be loaded within 20% of the task time in order to be shipped.
    e. None of the above.

15. What is designated by a "1" in the M column?
    a. That the order is a pickup order.
    b. That the order is a regular order.
    c. That the order is a high-priority order.
    d. That the order is a rush order.
    e. None of the above.

16. Which of the following is true in this task?
    a. If you attempt to ship a truck that is overloaded by more than 20%, a penalty is assessed.
    b. If you attempt to ship a truck to fewer than three zones, a penalty is assessed.
    c. If you attempt to rush deliver the regular orders, a penalty is assessed.
    d. If you attempt accept more than 20 orders, a penalty is assessed.
    e. None of the above.

17. Which of the following is true in this task?
    a. Rush orders must be shipped within two (2) days.
    b. Regular orders must be shipped within two (2) days.
    c. Incoming orders (orders in queue) must be shipped within two (2) days.
    d. Pickup orders must be shipped within two (2) days.
    e. None of the above.

18. What key or keys allow you to ship a pickup order?
    a. CTRL-Y.
    b. CTRL-A.
    c. CTRL-P.
    d. CTRL S.
    e. None of the above.

19 How is a regular order designated?
    a. By a "2" in the M column.
    b. By a "1" in the M column.
    c. By a "0" in the M column.
    d. By a "P" in the M column.
    e. None of the above.

20. One day is:
    a. The time limit placed on shipping rush orders.
    b. The time limit placed on shipping regular orders.
    c. The time limit placed on shipping accepted orders.
    d. The time limit placed on shipping incoming orders.
    e. None of the above.

21. A penalty is assessed if you:
    a. Attempt to ship pickup orders via the pickup window.
    b. Attempt to ship pickup orders on a truck.
    c. Attempt to accept more than 20 orders.
    d. Attempt to select orders for fewer than three zones.
    e. None of the above.

22. Which of the following is true in this task?
    a. Orders accepted for processing must be accepted into the top-right window.
    b. Orders accepted for processing must be shipped by the end of task.
    c. Orders accepted for processing must be sent to the pickup window.
    d. Orders accepted for processing must be assigned to a truck.
    e. None of the above.

23. Which of the following is true in this task?
    a. If you fail to ship all orders as soon as they are accepted, a penalty is assessed.
    b. If you fail to ship all orders within one (1) day, a penalty is assessed.
    c. If you fail to ship all regular orders within one (1) day, a penalty is assessed.
    d. If you fail to ship all orders on schedule, a penalty is assessed.
    e. None of the above.

24. How is a rush order designated?
    a. By a "P" in the M column.
    b. By a "0" in the M column.
    c. By a "1" in the Zone column.

d. By a "2" in the Zone column.
e. None of the above.

25. What is accomplished by pressing the ENTER key?
a. You can assign orders to a specific truck or the pickup window.
b. You can move between the top and bottom windows.
c. You can accept waiting orders for processing.
d. You can view truck capacity information.
e. None of the above.

26. Three is:
a. The minimum number of zones that a truck may deliver orders to.
b. The maximum number of zones that a truck may deliver orders to.
c. The maximum number of zones that a truck may deliver pickup orders to.
d. The maximum number of times that a truck may deliver orders during the task.
e. None of the above.

27. What is designated by a "P" in the processing (bottom, right-hand) window.
a. The location for assignment of truck capacity.
b. The location for assignment of orders to a specific truck.
c. The location for assignment of a pickup order.
d. The location for assignment of incoming orders.
e. None of the above.

28. What key or keys allow you to ship a truck?
a. CTRL-P.
b. CTRL-A.
c. CTRL-X.
d. CTRL S.
e. None of the above.

29. What is accomplished by pressing CTRL-A?
a. You can ship a truck.
b. You can ship the pickup order(s).
c. You can assign orders to a specific truck.
d. You can send pickup orders to the pickup window.
e. None of the above.

30. What key or keys allow you to ship a pickup order?
a. CTRL-Y.
b. CTRL-A.
c. CTRL-P.
d. CTRL S.
e. None of the above.

Appendix B

MVPDisUnd Learning Test B

Participant Number: __________ Date: ________________

Directions: Please circle the letter that corresponds to the answer that best completes the question.

IMPORTANT!!! Please DO NOT SKIP ANY QUESTIONS or leave any questions unanswered.

1. Order activation, identification, & assignment are:
    a. The activities that are required for successful task verification.
    b. The activities that are required for successful task identification.
    c. The activities that are required for successful order packaging.
    d. The activities that are required for successful order identification.
    e. None of the above is true in this task.

2. When the task begins, the first activity you may do is:
    a. Press CNTRL-A.
    b. Press CNTRL-Y.
    c. Use the arrow and ENTER keys.
    d. You may do either a. or c. at this time.
    e. None of the above is the correct activity.

3. When the task begins, you may:
    a. Accept, select, and/or package orders for shipping.
    b. b. Select, verify, and/or package orders for shipping.
    c. Activate, identify, and/or assign orders for shipping.
    d. Both a and b are true in this task.
    e. None of the above is true in this task.

4. If orders are not processed quickly, what will occur?
    a. Penalties may be assigned because of delay in processing orders.
    b. Penalties may be assigned because you correctly followed task rules.
    c. Your overall score will be decreased because of penalties.
    d. Both a and c are true in this task.
    e. Only a is true in this task.

5. To activate waiting orders (orders in queue), you must:
    a. Determine that there is room for more orders in the bottom, left-hand window.
    b. Determine that there is room for more orders in the bottom. right-hand window.
    c. Have space for more orders in the top, right-hand window.
    d. Both a and b are true in this task.
    e. None of the above is true in this task.

6. When you accept waiting orders, you are:
    a. Selecting those orders for a truck.
    b. Activating those orders for processing.
    c. Accepting those orders for the pickup window.
    d. Shipping those orders on a truck or via the pickup window.
    e. None of the above is true in this task.

7. To activate waiting orders (orders in queue), you must:
    a. Have space for more orders in the top, left-hand window.
    b. Have space for more orders in the bottom. left-hand window.
    c. Have space for more orders in the top, right-hand window.
    d. You may not activate waiting orders at any time during the task.
    e. None of the above is true in this task.

8. To activate waiting orders (orders in queue), you must:
    a. Have space for more orders in the bottom, left-hand window.
    b. Have space for more orders in the top, right-hand window.
    c. Have space for more orders in the bottom, right-hand window.
    d. You may activate waiting orders at any time during the task.
    e. None of the above is true in this task.

9. What does a "1" in the M column designate?
    a. That the order is a rush order.
    b. That the order is a regular order.
    c. That the order is a pickup order.
    d. That the order should be sent to zone 1.
    e. None of the above is true in this task.

10. To determine if an order is a rush order, you should first:
    a. Look for a "1" in the M column.
    b. Look for a "2" in the M column.
    c. Look for a "3" in the M column.
    d. Look for a "0" in the M column.
    e. None of the above is the correct activity.

11. How do you determine that an order is not a partial order?
    a. By looking at the order's invoice number(s).
    b. By looking at the order's identification number(s).
    c. By looking at the order's schedule number(s).
    d. By looking at the information in the M column.
    e. None of the above is the correct activity.

12. What does an order's zone number determine?
    a. Where to ship the order by truck.
    b. The order's delivery schedule for delivery.
    c. How soon the order must be shipped.
    d. When to send the order to the pickup window.
    e. None of the above is the correct activity.

13. To determine if an order should be sent to the pickup window, you should first:
    a. Look for a "3" in the M column.
    b. Look for a "0" in the M column.
    c. Look for a "2" in the M column.
    d. Look for a "1" in the M column.
    e. None of the above is the correct activity.

14. How do you know that an order should be sent to the pickup window?
    a. There is a "1" in the zone column for the order.
    b. There is a "1" in the M column for the order.
    c. The order has been accepted for processing.
    d. There is a "5" in the QTY column for the order.
    e. None of the above is true in this task.

15. To ship a pickup order:
    a. Press ENTER.
    b. Press CNTRL-A.
    c. Press F1.
    d. Press CNTRL-Y.
    e. None of the above is the correct activity.

16. For pickup orders, you can use the arrow and ENTER keys to:
    a. Select and assign the order to a truck.
    b. Select and assign the order to the pickup window.
    c. Ship the order via the pickup window.
    d. Ship the order according to zone.
    e. None of the above is true in this task.

17. In order to ship an order by truck, you must first:
    a. Determine that a truck is available for a pickup order.
    b. Determine that a truck is loaded to its capacity.
    c. Determine that a truck is currently delivering orders.
    d. Determine that there is room in the queue.
    e. None of the above is the correct activity.

18. To determine if an order may be shipped with other orders, you should:
    a. Look at the information in the M column.
    b. Look at the order's zone number and invoice number(s).
    c. Look at the number of orders in queue.

d. Look at the order's truck capacity and weight information.
e. None of the above is the correct activity.

19. After assigning orders to a truck, you may:
    a. Select orders according to invoice number(s).
    b. Assign those orders to the pickup window.
    c. Assign different orders to another truck.
    d. Accept those orders into the active (processing) window.
    e. None of the above is true in this task.

20. After an order has been assigned to a truck, you may:
    a. Continue selecting orders for the same truck.
    b. Ship the truck.
    c. Unload the pickup window.
    d. Both a. and b. are possible.
    e. None of the above is the correct activity.

21. If you attempt to ship a truck that is overloaded:
    a. The truck may go out, but the orders remain in the window.
    b. The truck may go out, but the orders remain in the queue.
    c. The truck may go out, but a penalty is assessed.
    d. The truck will not go out, and a penalty is assessed..
    e. None of the above is true in this task.

22. Immediately after shipping a truck, you should:
    a. Accept new orders for processing for processing.
    b. Ship new orders.
    c. Assign orders to a specific truck(s).
    d. You may do either a. or c. at this time.
    e. None of the above is the correct activity.

23. Why would you remove orders from a truck?
    a. To avoid getting a penalty because the truck is underloaded.
    b. To avoid getting too many orders in the queue.
    c. To get more orders from the queue.
    d. To avoid a penalty for having too many orders in the queue.
    e. None of the above is true in this task.

24. If truckload exceeds truck capacity, you should:
    a. Remove excess pieces according to invoice number & QTY information.
    b. Remove excess units according to zone number & QTY information.
    c. Both a and b are correct.
    d. Assign more orders to the truck.
    e. None of the above is the correct activity.

25. Immediately after shipping a truck, you should:
    a. Press CNTRL-A.
    b. Press CNTRL-P.
    c. Use the arrow and ENTER keys.
    d. You may do either a. or c. at this time.
    e. None of the above is the correct activity.

26. How do you determine which order(s) should be loaded onto which truck?
    a. By looking at the order's zone number.
    b. By looking at the M column.
    c. By looking at the order's delivery schedule.
    d. By looking at the order's capacity number.
    e. None of the above is true in this task.

27. After an order has been assigned to the pickup window, you may:
    a. Use the CNTRL-Y key(s) to assign the order to a truck.
    b. Use the arrow and ENTER keys to select another pickup order.
    c. . Use the number and ENTER keys to select another order by zone.
    d. Use the CNTRL-Y key(s) to ship the order on a truck.
    e. None of the above is the correct activity.

28. How do you determine the number of pieces in an order?
    a. By looking at the information in the QTY column.
    b. By looking at the order's invoice number(s).
    c. By looking at the information in the zone column.
    d. By looking at the information in the M column.
    e. None of the above is the correct activity.

29. After an order has been assigned to the pickup window, you may:
    a. Press CNTRL-A to ship the pickup order.
    b. Press CNTRL-Y to ship the pickup order.
    c. Press CNTRL-S to ship the pickup order.
    d. Press ENTER to ship the pickup order.
    e. None of the above is the correct activity.

30. If orders are not processed accurately, what will occur?
    a. Penalties may be assigned because of rule violation(s).
    b. Your overall score will be decreased because of penalties.
    c. Both a and b are true in this task.
    d. Penalties may be assigned because you correctly followed task rules.
    e. Both a and d are true in this task.

Appendix C
Test Perceptions

Participant Number: __________ Date: ________________

Directions for Survey:
Please read each of the following statements. Rate the extent to which you agree or disagree with each statement. (Enter the appropriate number on the line before each question.) Even if you are unsure of an item, please answer it anyway.

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

____1. I believe that the test I just completed was designed to measure my general intelligence.

____2. I believe that the test I just completed was designed to measure my knowledge of a particular task.

____3. I do not understand what the test had to do with this job.

____4. I can not see any relationship between the test and what I think is required by the job tasks.

____5. It would be obvious to anyone that this test is related to the job tasks.

____6. The actual content of this test was clearly similar to the job tasks listed above.

____7. Doing well on this test was important to me.

____8. I didn't put much effort into this test.

____9. I tried my best on this test.

____10. While taking this test, I concentrated and tried to do well.

____11. I just didn't care how I did on this test.

____12. I want to be among the top scorers on this test.

____13. I was extremely motivated to do well on this test.

Appendix D
Demographics Survey

Participant Number: __________ Date: ___________

Directions for Survey:
Please read and answer each of the following questions. (Please circle the appropriate letter.) Even if you are unsure of an item, please answer it anyway.

1. Age ________ (Please enter number of years only.)

2. Race

a. African American

b. Asian American

c. Hispanic American

d. White American

e. Other (Please identify) _________________

3. Gender:

a) Female b) Male

4. College Ranking

a) Freshman b) Sophomore c) Junior d) Senior

5. College Major

a) Business b) Communications c) Computers d) Education

e) Engineering f) Mathematics g) Psychology h) Sociology

i) Other