PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

CHANGING GEARS:
MODELING GENDER DIFFERENCES IN PERFORMANCE
ON TESTS OF MECHANICAL COMPREHENSION

By

James A. Grand

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Psychology

2008

ABSTRACT

CHANGING GEARS:
MODELING GENDER DIFFERENCES IN PERFORMANCE
ON TESTS OF MECHANICAL COMPREHENSION

By

James A. Grand

For over 60 years, the selection, training and vocational coaching assessments of individuals in mechanically-inclined occupations has often involved the administration of mechanical ability tests. However, men have consistently outperformed women on such tests, creating the possibility for unfair hiring or human resource practices in these situations. Unfortunately, the causes of these achievement differences are still largely unknown. As such, this study proposes and examines a functional gender differences model that examines predictors of mechanical ability test performance beyond gender that could potentially be leveraged to diminish the substantial performance differences found in mechanical and other similarly gender-biased cognitive ability tests. Specifically, the model investigates one's gender role identification, gender stereotype endorsement, mechanical interests/experiences, and mechanical self-efficacy as influences of mechanical comprehension. The results of the study revealed that, even when controlling for gender, an individual's self-efficacy for accomplishing mechanically-related activities was strongly and positively related to test performance; furthermore, self-efficacy was predicted by one's mechanical interests and experiences, which were meaningfully related to gender role identification. Practical implications and limitations of the model's results in terms of capturing gender differences in cognitive ability tests and topics for future related research are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

CHANGING GEARS:
MODELING GENDER DIFFERENCES IN PERFORMANCE ON TESTS OF MECHANICAL COMPREHENSION

INTRODUCTION

I was born a mechanic, and made a barrel before I was ten years old. The cooper told my father, "Fanny made that barrel, and has done it quicker and better than any boy I have had after six months' training." My father looked at it and said, "What a pity that you were not born a boy so that you could be good for something. Run into the house, child, and go to knitting."

— Frances D. Gage (Stanton, Anthony, & Gage, 1882)

The opening quotation above from American suffragist Frances Gage succinctly and strikingly highlights an issue that has long stirred debate in court rooms, board rooms, class rooms and living rooms the world over—are men and women inherently different in terms of what they can and can not do or are such gender differences only a reflection of the environment in which we live? Such a definitive and encompassing answer to this question is not to be found in the current research, nor is such a resolution likely to be found in any research in the near-future as well. But, by limiting the scope of broad gender differences in aptitude to a more narrow range of dimensions, it may be possible to incrementally build a better understanding of the processes through which such differences arise and are maintained. As such, the present study's focus is concerned with understanding the differences observed between males and females in the specific cognitive ability of mechanical comprehension.

Though much more will be relayed on this topic in the sections to follow, empirical research has long noted the superiority of males over females on cognitive tests of mechanical comprehension (e.g., Bennett & Cruikshank, 1942). This finding is particularly troubling given that such tests are used for a wide variety of industry practices in mechanically related jobs, including selection, training assessment and

1

vocational counseling (Super & Crites, 1962). The severity of the problem is amplified even further when the status of the current labor market is brought into consideration. Within the last twenty years, the number of women who have begun to move into traditionally male dominated work has increased dramatically (Blau & Hendricks, 1979; England, 1981; Beller, 1982), including jobs for which tests of mechanical comprehension are typically administered as a means of selection (Super & Crites, 1962). Though such mechanical jobs (such as production and maintenance work) are seldom considered "high-status," they often offer better pay and more steady employment than comparative female dominated occupations (such as clerical work, Muchinsky, 2004). Thus if females truly are deficient in mechanical comprehension, administering tests for this construct as part of a selection system will almost certainly prevent women's entry into these potentially more attractive mechanical jobs; however, if the observed male-female difference is instead a response to some sociocultural phenomenon (i.e., stereotypes, gender roles, etc.), it may be possible to reduce the performance gap through training or some other means of education.

The purpose of this research is to propose and test a model of the antecedents related to the performance disparity across gender found on tests of mechanical comprehension (also referred to as mechanical aptitude or ability). The development of the components and linkages of the model will be established through an overview of previous research in the areas of mechanical comprehension, gender differences, and test performance. Following this review, the study's research hypotheses are given and discussed in relation to the gender model and the relevant literature, and the methodological procedures undertaken to test these hypotheses are described. Finally, an

analysis of the data is presented and interpretations of the study's conclusions are drawn for the purposes of aiding future research and development in the area of gender differences in mechanical ability testing specifically, and hopefully cognitive ability testing in general.

Definition of the Mechanical Comprehension Construct

It is perhaps helpful to approach the construct of mechanical comprehension through the framework of Spearman's (1927) Two-Factor theory of intelligence. According to the theory, a person's cognitive functioning can be broken down and analyzed into two distinct factors—a general (*g*) factor and specific (*s*) factors. As the name implies, *g* is an overall mental capacity or intellectual capability that operates in all forms of cognitive functioning. *s* is similarly defined, but rather than operating across many areas of cognitive functioning, it is more narrowly restricted in its range. With the development of more sophisticated factor analysis techniques (Guilford, 1948; Thurstone, 1948), researchers soon after modified the Two-Factor theory of intelligence to include lower- and higher-order factors that lie in between *g* and *s* in terms of their influence on the performance of certain abilities.

Vernon (1950) combined these findings to create the hierarchical structure of human abilities, which demonstrated how *g* could be usefully conceptualized as the highest-order factor among more specific levels of intellectual functioning. The model extended from the various permutations of *s* on up through successively more encompassing factors (see Figure 1). Through factor analysis of *g* ability tests administered in selective military populations, Vernon discovered that two major group
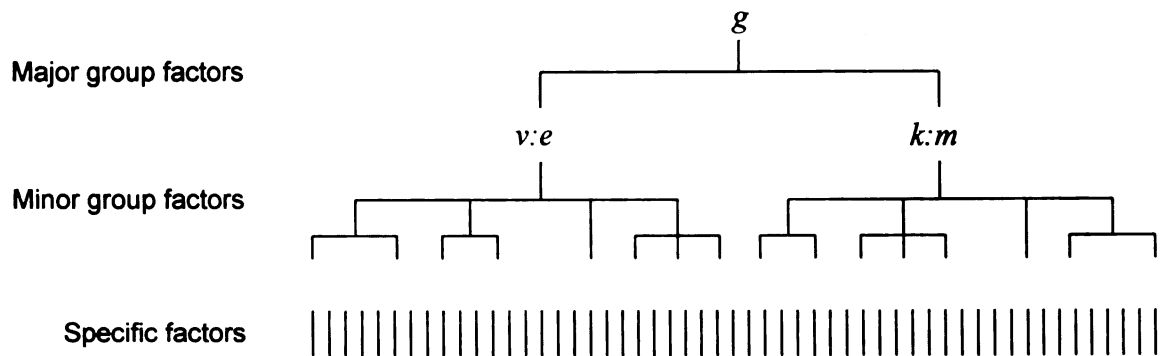
Figure 1. Diagram of hierarchical structure of human abilities (Vernon, 1950).

factors reliably emerged. The first of these he labeled a verbal-numerical-educational

factor (represented as *v:ed*), and the other a practical-mechanical-spatial-physical factor

(*k:m*). Given sufficiently specific tests, Vernon states, each of these types further

subdivides into a series of minor group factors. In the case of the *k:m* factor, these

include spatial, manual and mechanical information subfactors, among others. Based on

Vernon's hierarchical structure, then, it can be concluded that mechanical comprehension

does indeed appear to load on both *g* and spatial/perceptual factors (Guilford, 1947).

Indeed, empirical work seems to support this theory as Bennett (1969) reports that early

versions of the Bennett Mechanical Comprehension Test (BMCT) demonstrated

correlations between .40 and .60 with various intelligence and spatial ability tests.

Nevertheless, because of the high face validity, reliability, and predictive validity for tests

of mechanical comprehension among jobs requiring the aptitude, the construct has gained

considerable support as one of the most accurate cognitive assessments of performance in

mechanically-inclined occupations (Muchinsky, 2004).

As demonstrated by Vernon (1950), a number of cognitive abilities are closely

related to mechanical comprehension; thus it is necessary to precisely identify the

construct for the purposes of the current research study. To this end, mechanical

4

comprehension will herein be defined as the ability to learn about, perceive and understand the operation of common physical principles and mechanical elements in practical situations, in both an implicit and explicit manner (Bennett, 1969; Weisen, 1999). This definition was specifically chosen to emphasize the natural, instinctive and intuitive fashion through which mechanical comprehension is experienced, as opposed to any proficiency gained through formal instruction in physics or mechanical diagnostics. In addition, it is intended to exclude any other closely related cognitive factors, such as spatial visualization (Wittenborn, 1945; Guilford, 1948) or perceptual speed and acuity (Super & Crites, 1962).

Early Research on Mechanical Comprehension

Spurred by the widespread acceptance of cognitive ability testing during the post-World War I/pre-World War II era, researchers began to take a vested interest in the study of mechanical aptitude in the early 1900s. Although tests of mechanical aptitude had been in circulation for nearly half a decade (Stenquist, 1923), the first empirical studies of mechanical comprehension were not published until 1928 by Cox and, shortly thereafter, Patterson, Elliot, Anderson, Toops, and Heidbrieder in 1930.

Following Spearman's (1927) Two-Factor theory of intelligence, Cox (1928) set out to determine whether mechanical aptitude was a distinguishable construct apart from *g*. Through a series of paper-and-pencil tests, Cox attempted to isolate and capture the unique cognitive mechanisms involved in mechanical comprehension. In order to minimize any potential construct contamination, the tests were designed such that any instruction in physics or formal mechanics was not necessary or specifically useful in answering the questions. For example, the Test of Mechanical Models presented

participants with the beginning and end states of a particular mechanical operation from which they were to derive the intermediary processes. Based on factor analysis of his data and a number of divergent validity studies, Cox concluded that there was indeed a cognitive ability distinct from *g* that predicted performance on tests of mechanical comprehension which he termed mechanical aptitude. Although much later, researchers have suggested that Cox actually captured the closely related spatial visualization factor rather than a distinctive mechanical comprehension dimension in this initial study (Super & Crites, 1962); nevertheless, these efforts generated a great deal of early empirical interest in the area of mechanical aptitude.

Unlike the strictly theoretical approach adopted by Cox (1928), Patterson et al. (1930) set out to develop a practical measure that could be used to identify individuals who displayed a proficiency in mechanical comprehension. According to Patterson et al.'s definition, mechanical ability is taken "to refer to whatever capacities and abilities are necessary for certain kinds of work—specifically, work that involves the manipulation of tools, the operation of machinery, and the planning and execution of pieces of work which involve these and similar activities" (p. 5). Rather than create a unique bank of tests for these purposes, the researchers revised and selected items from a number of previously published tests including the Minnesota Assembly test, Minnesota Spatial Relations test, and Minnesota Paper Form Board test, among others. Based on an extensive criterion validity effort, a weighted average scoring system was derived to achieve a final, unique score for mechanical ability. Patterson et al. administered their final test battery to a large number of junior high school boys, in the process establishing

a large normative dataset, which would serve as the standard for future tests of mechanical comprehension.

In the years following Cox (1928) and Patterson et al.'s (1930) inquiries, tests of mechanical comprehension were readily adopted into industry practice for a variety of purposes, including admission into training programs (Martin, 1951) and trade schools (Patterson, 1956), vocational counseling (Cantoni, 1955; Klare, Gustafson, Mabry & Shuford, 1955), and, especially, as instruments of selection (Harrell & Faubion, 1940; Ghiselli & Brown, 1951; Wolff & North, 1951). In addition, mechanical aptitude testing has been applied extensively in the military arena for over 60 years as a means of enlistment and duty assignment for identifying individuals qualified for service in a particular branch. Even as early as 1947, Guilford and Lacey found that the BMCT—which to this day remains the most widely used test of mechanical comprehension (Muchinsky, 2004)—was one of the top predictors of performance by military fighter pilots during World War II. In fact, tests of mechanical comprehension were widely used during World War II, so much so that "...the War Department considered the measurement of mechanical ability of such importance that a Mechanical Aptitude test was administered at reception centers to all enlisted men who could read or write" (Tiffin, Knight & Asher, 1946, p. 236). It would appear, then, that the construct of mechanical comprehension has had a very real impact from both an empirical and practical perspective.

Empirically Observed Gender Differences in Mechanical Comprehension

The measurement and identification of mechanical comprehension has had a fairly long history of research and industry use—however, very few advances in terms of

our understanding of the construct and our ability to accurately assess it have been made in recent years (Stumpf, 1995). Muchinsky (2004) suggests that this may be a result of the "'unfashionable' nature of studying industrial production workers" (p. 23) or perhaps because of its relatively restricted area of applicability in today's workforce. Whatever the case may be, the lack of progress is unfortunate, especially given that mechanical aptitude is one of the most reliable and consistent cognitive differences across genders (Bennett & Cruikshank, 1942; Antill & Cunningham, 1982; Anastasi, 1988; Stumpf, 1995; Halpern, 2000).

A number of studies have observed the marked gender difference in performance on tests of mechanical comprehension. In assessing a variety of standardized aptitude tests, Stanley, Benbow, Brody, Dauber, and Lupkowski (1992) noted significant gender effect sizes on tests of mechanical ability, with d coefficients ranging from .66 to .89. Bennett and Cruikshank (1942) also reported large gender differences in performance when conducting validation studies on an earlier version of the BMCT. An approximately 15 point mean difference in test scores was observed between males and females in Grades 10 through 13; using somewhat older age groups (applicants to nursing schools, $M = 23.37$, versus applicants for positions as firemen and policemen, $M = 36.78$) resulted in only a very slight improvement in the performance gap. Not surprisingly, the authors found gender to be a significant predictor of performance on the test, explaining from 49% to 79% of the variance in test scores across the different samples.

A somewhat more evolved pattern of gender differences in mechanical comprehension test performance was found using extensive validation data collected for the Mechanical Reasoning subscale of the Differential Aptitudes Test (DAT). Bennett

(1969) reported that while the mean within-group test scores for both males and females increased each year between Grades 8 and 12, the mean difference between groups (i.e. the performance gap) also increased each year. Thus, although males and females demonstrated improved scores on the test from one year to the next, the improvement by males outpaced females such that the difference in performance was actually much greater at older ages.

Feingold (1988) further expounded these results by analyzing approximately 30 years of normative data collected on the eight subtests of the DAT for gender differences in test performance. The portion of Feingold's results concerning mechanical comprehension is reproduced in Table 1. As can be seen from the table, males once again significantly outperformed females on the Mechanical Reasoning subtest at every grade level and at every year. Though not apparent from the limited data shown in Table 1, the magnitude of the gender differences in performance on Mechanical Reasoning were actually the largest of any of the eight DAT subscales. The analysis by Feingold also revealed a significant age effect, such that the performance gap on tests of mechanical comprehension between males and females grew larger at each successive age group. The findings by Bennett (1969) and Feingold (1988) that the performance gap on tests of mechanical comprehension widens as a function of age is a common one in the literature on developmental gender differences in specific cognitive abilities (Maccoby & Jacklin, 1974); however, mechanical aptitude appears to be one of the most intensely pronounced (Anastasi, 1981; 1988).

As an aside, the mechanical comprehension data from the Feingold (1988) study also demonstrates an interesting "secular trend" that has been noted by many researchers

9

Table 1

Cognitive Gender Differences (ds) on the Mechanical Reasoning Subtest of the DAT by Grade and Year of Standardization

| Year | Grade in School | | | | | Year M |
	8	9	10	11	12	
1947	1.02	1.24	1.34	1.48	1.55	1.33
1962	.80	.90	1.00	1.06	1.23	1.00
1972	.74	.79	.84	.88	.92	.83
1980	.66	.68	.75	.82	.89	.76
Grade M	.80	.90	.98	1.06	1.15	.98

Positive *d* indicates higher male performance.
Table reproduced from Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*(2), 95-103.

studying longitudinal gender differences in specific cognitive ability (Stumpf, 1995).

The secular trend refers to the shrinking performance gap in specific cognitive abilities

between male-female cohorts (represented by examining the columns in Table 1 from top

to bottom), as opposed to the increasing performance gap *within* male-female cohorts

(represented by examining the rows in Table 1 from left to right). In large part, there is

empirical uncertainty as to why such a pattern has been observed. Flynn (1984) reported

over two decades ago that the average American's IQ (a general measure of *g*) increased

by 12 points between 1932 and 1972, thus hypothesizing that there may be a generational

effect on cognitive ability (however, this "Flynn Effect" has not been universally

replicated, see Sundet, Barlaug, & Torjussen, 2004; additionally, there is at least some

evidence stating that the growth has slowed in recent years, Teasdale & Owen, 2000).

Extrapolating from this hypothesis, the hierarchical structure of human abilities

(Vernon, 1950) suggests that an increase in *g* should positively correlate with increases in

particular specific cognitive abilities as well. Therefore, it could be argued that gains in *g*

for women have contributed to greater improvements in mechanical ability while

manifesting in different areas of functioning for men. At a more universal level of explanation, it is likely that temporally-related changes in the brain/biology and sociological characteristics (see section below on *Theories of Gender Differences*) of the general population have contributed significantly to the observed secular trend as well (Teasdale & Owen, 2000). Of course, these speculations are hypothetical, and the causes and explanations for this trend are still unclear in the research literature; however, despite the shrinking performance gap in mechanical comprehension, the magnitude of the male-female differences that do still exist are quite large and thus warrant further study.

Theories of Gender Differences in Mechanical Comprehension

At the definite risk of oversimplification, there are two broad classes of explanations that attempt to capture how and why gender differences in aptitudes arise: biological factors and direct socialization influences from the environment (Maccoby & Jacklin, 1974, present an extensive review of much of the early research conducted in these areas and serves as an excellent point of entry into the discussion). However, it is important to note that these theories are not mutually exclusive, and they should not be considered as such. It is simply convenient to separate and study both influences as though they were distinct in order to better isolate the underlying phenomena captured by each. There is little argument that the interaction between nature, nurture and the various psychological processes that accompany them is the most accurate description of the underlying differences that exist between men and women (Wood & Eagly, 2002). But, as is often the case, the requirements and limitations of empirical research outweigh our ability to wholly capture the true nature of the event and thus we must rely on a less desirable, "piecemeal" approach. Nevertheless, a great deal of research effort has gone

into explaining the origins and sustaining forces of gender differences in specific cognitive abilities, of which only a cursory glimpse will be presented below.

Biological factors. The argument for biological differences between the genders accounting for differential cognitive performance has been approached in a number of different ways and has implicated a number of possibilities. One of the most widely studied areas in this arena concerns the effect of hormones on cognitive ability (e.g., Rogers, 1976; Messent, 1976; Harris, 1978; 1981). The basic notion follows that during crucial brain development stages, males and females are exposed to and come to possess differing amounts of the primary sex hormones, including estrogen, progesterone and testosterone, among others (Hamburg & Lunde, 1966; Halpern, 2000). Based on the relative proportion and activation of these hormones, areas of the brain develop differently—thus, as there should typically be less within-sex variability than between-sex variability in terms of relative levels of the sex hormones, the brains of males and females should tend to develop in different fashions thereby resulting in structural brain differences across the sexes that could account for variability in performance on certain tasks.

For example, one of the most influential theories concerning the impact of hormones on cognitive ability was put forth by Geschwind and Galaburda (1987). In this theory, it was proposed that higher concentrations of prenatal testosterone slow the growth of neuronal connections within the brain's left hemisphere. This often results in right hemisphere dominance (also known as brain lateralization), in which the right hemisphere of the brain has much more finely tuned control and "processing power" than the left hemisphere. As a normally developing male will be exposed to higher levels of

prenatal testosterone than a normally developing female (because testosterone is naturally produced by the male testes), males are much more likely to be right-brain dominant, a finding which has been reliably reproduced by many researchers (cf. Halpern, 2000). On the other hand, females are more likely to exhibit bilateral brain functioning (equal activation in both hemispheres) and are, consequentially, much more likely to employ both hemispheres when approaching a cognitive task (Harris, 1981).

Thus, Geschwind and Galaburda's (1987) theory predicts that males should be more efficient at tasks that require primarily right-brain activation while being less proficient at left-brain tasks—and indeed, there has been some research to support this claim. Mathematical reasoning and spatial ability skills represent two notable cognitive abilities that have been demonstrated to be right-brain dependent, and two notable areas in which males have typically outperformed females (Maccoby & Jacklin, 1974; Harris, 1981; Halpern, 2000). In addition, males have also been found to possess the majority of the language production and reading problems (such as dyslexia) normally associated with active left-brain functioning. Furthermore, these cognitive verbal abilities are both areas in which females have demonstrated markedly better performance than males (Maccoby & Jacklin, 1974; Halpern, 2000). Based on these findings then, and recalling the earlier discussion of the relationship between mechanical comprehension and spatial ability (Vernon, 1950), there is at least indirect evidence to support the notion that the differences in performance found on tests of mechanical comprehension can partly be attributed to the differential brain functioning between males and females.

Gender socialization. The quotation from Frances Gage presented at the beginning of this paper provides a blunt, yet poignant, example of the gender

socialization process. Theories of gender socialization generally posit that, over the course of the natural lifespan and especially during the influential childhood years, males and females are differentially treated by others in their immediate environment as a result of established sociocultural norms. Such treatment manifests in the disparate types of opportunities, reinforcement/punishment, and training that are made available to the members of each sex. As a result of these experiential differences then, males and females come to acquire and gain proficiency in dissimilar sets of skills and aptitudes (Maccoby & Jacklin, 1974; Jacklin & Reynolds, 1993; Lott & Maluso, 1993; Wood & Eagly, 2002). Perhaps put more succinctly, gender socialization is the process by which individuals come to learn and adopt the socially created and accepted norms, goals and values of their gender (Eccles, 1987). Since these gender-appropriate goals and values are typically quite different, men and women often become better or more skilled with certain tasks and abilities than members of the opposite sex (Stumpf, 1995; Halpern, 2000).

There are a number of different theories that overlap with or fall under the umbrella of "gender socialization." For example, social learning theory (Bandura, 1969), though typically treated as a more general explanation for the behavioral adoption patterns seen in children, is very similar to the above treatment of gender socialization. In social learning theory, children are reinforced for certain desired behaviors by the social agents in their environment. In addition, social learning theory also emphasizes the importance of behavior modeling, in which children choose to imitate the behaviors of the influential others around them. From the perspective of gender then, social learning theory predicts that children choose to imitate models of the same sex, and thus

acquire the behaviors and associated skills and abilities displayed by their male or female role models.

Also similar to the gender socialization hypothesis is gender schema theory, which proposes that individuals develop a systematic, formulated set of ideas about what is and what is not male or female based on the information available to them in the environment (Bem, 1981). This network of associations (which incorporates a diverse number of categories including perceptions about affect, attitudes, objects, behaviors, etc.) then becomes a "cognitive filter" through which any new information and experiences pertaining to men and women are passed through, organized, and then adapted into the previous gender schema. To this end, one's gender schema is constantly changing and evolving to incorporate and process new information pertaining to the roles of males and females. One of the more important implications of gender schema theory is that individuals learn to "sort people, behavior, and attributes into the culture's definitions of feminine and masculine" (Jacklin & Reynolds, 1993, p. 201). It logically follows, then, that individuals would also attempt to place themselves within their gender schema based on their own behaviors, attitudes, etc. Therefore, those people who are "gender schematic" and adhere rigidly to their associations pertaining to gender roles are more likely to adhere to the sociocultural norms about male-female differences in skills and aptitudes than individuals who are "gender aschematic" and not as reliant on gender-role associations (Bem, 1985).

In terms of the impact of gender socialization on the development of mechanical comprehension, a return to Frances Gage's short-lived experience as a mechanic's apprentice depicted earlier yet again offers an exceptionally concrete and vivid means of

explanation. Quite clearly, Gage believes and definitively states that she has always had some innate proficiency with all things mechanical, and has even demonstrated that she can perform a mechanical task as well or better than any similarly trained male. However, even despite her obvious aptitude, her father wholly discourages the pursuit of such a traditionally masculine activity and instead instructs her to return to the house and take up a more "appropriate" feminine activity. Presumably, once "in the home" and in the presence of other same sex feminine social agents, Gage is not likely to find an influential model that she could imitate or learn from in order to build the physical skills and cognitive abilities required of a mechanic. Thus, she is denied the opportunity to pursue further training and is shut out from additional experiences that might have served to improve her level of mechanical comprehension because of a sociocultural norm.

Proposed Model of Gender Differences in Mechanical Comprehension

The functional model of gender differences in mechanical comprehension that will be tested by the current study is presented in Figure 2. Though the major variables and hypotheses concerning their linkage will be discussed and justified in the sections to follow, a general overview of the model serves as a useful starting point. The basic organization follows a direct mediation structure in which self-efficacy for mechanical tasks is proposed to influence the degree to which an individual's level of mechanical interest, knowledge, and experience predicts performance on a test of mechanical comprehension. Two additional variables are also introduced which are believed to have an impact on the direct mediating linkage. The first, gender role identification, is predicted to influence the level of mechanical interest, knowledge, and experience
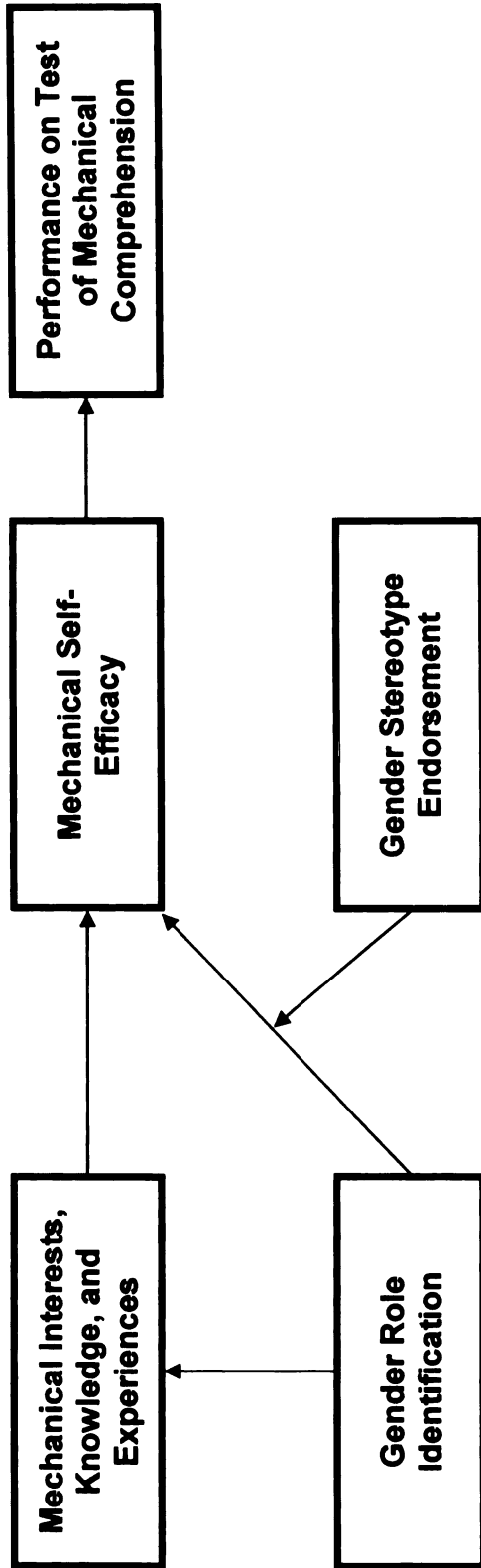
Figure 2. Proposed model of gender differences in mechanical comprehension as measured by test performance.

reported by an individual. The second, endorsement of gender stereotypes, is predicted to moderate the link between mechanical interest, knowledge, and experience and mechanical self-efficacy. The overall model can be observed for both males and females to identify where inconsistencies in the linkages lie.

The model proposed by the current study offers a significant point of departure from previous explanations of gender differences in specific cognitive ability in that it focuses more attention on the individual-level, process-outcome links that result in differential aptitude testing scores. As the previous section on theories of gender differences outlined, a great deal of theoretical and empirical effort has gone towards identifying the broad antecedents that account for *group-level* gender differences—but far less consideration has been given to the interaction between gender differences and *individual differences* (i.e. mechanical experiences, mechanical self-efficacy, etc.) that might impact the measurement of specific aptitudes. The fact that gender plays an important role in the development and expression of certain cognitive abilities is not questioned by the current study; however, the direct gender-to-performance relationship seems an overly simplified explanation. Though the link that connects gender to aptitude testing performance has been well documented, there is still a significant amount of variance left to be explained (Antil & Cunningham, 1982). Understanding potential individual difference variables that contribute to the performance gap and how they interact would help to more precisely elucidate how variability in specific aptitudes plays out across gender and gender-related characteristics. Thus one of the goals of the present research is to validate a relational model that incorporates known outcomes of gender

differences with relevant individual differences in order to explain performance variation on tests of mechanical comprehension.

Mechanical interests, knowledge, and experiences. Bandura's (1977) early social cognitive theory proposed that feelings of self-efficacy arose primarily from four distinct sources: performance accomplishments (mastery), vicarious experience (modeling), verbal persuasion, and emotional arousal. While conceptually useful, these categories are fairly broad and not always helpful for identifying their direct influence on self-efficacy. To that end, Gist and Mitchell (1992) proposed, though never empirically tested, a restructuring of Bandura's social learning theory (1977) in which the determinants of self-efficacy were thought to lie in a 2 x 2 matrix composed of a dimension's variability within an individual and a person's "controllability of the causal influence" (p. 196). The variability of a dimension is conceptualized as high or low, depending on how stable the determinant is over time and situation, whereas the locus of the determinant can be either external or internal, based on whether the individual has control over the expression of the causal influence. As the external cues are proposed to be entirely a function of the task an individual is faced with (e.g., available resources, feedback, environmental distractions), they are less important for the purposes of the current study as they should be a constant across all participants. What should be less consistent, however, are the low variability/internal and high variability/internal determinants of self-efficacy. These include variables such as knowledge, skills, abilities, and interests, among others. Though no particular mention is made by Gist and Mitchell (1992) as to the degree of influence each of these variables has on self-efficacy, it should nevertheless be expected that that the direction of the relationship will be positive, such that:

Hypothesis 1: Individuals with more interest in, greater knowledge of, and more

experiences with mechanically-related subject matter will report higher ratings of

mechanical self-efficacy.

Mechanical self-efficacy. Self-efficacy can broadly be defined as "...people's

judgments of their capabilities to organize and execute courses of action required to attain

designated types of performances." (Bandura, 1986, p. 391). Based on the above, then, it

follows that an individual's level of self-efficacy should be highly predictive of a targeted

performance outcome—and indeed, self-efficacy has been demonstrated to be a reliable

predictor of performance across many different areas of functioning (e.g. Stajkovic &

Luthans, 1988; Mathieu, Martineau & Tannenbaum, 1993; Moritz, Feltz, Fahrbach &

Mack, 2000); however, the degree to which the relationship is meaningful depends in

large part on the level of specificity at which it is measured (Chen, Gully & Eden, 2001).

In the manner in which Bandura (1977; 1986) describes the construct, self-

efficacy is intended to be restricted to task-, domain- and situation-specific instances.

Thus, through performance accomplishments, vicarious experiences, verbal persuasion or

emotional arousal directed towards a *particular area of functioning*, an individual's self-

efficacy for performing domain-related tasks is believed to improve and ultimately

translate into improvements in actual performance. However, a number of researchers

have since pushed for the inclusion of general self-efficacy (GSE)—a singular, "trait-

like" evaluation of one's capability to perform across all areas—as an adequate predictor

of performance as well (e.g. Sherer et al., 1982; Tipton & Worthington, 1984; Woodruff

& Cashman, 1993; Chen, Gully, Whiteman & Kilcullen, 2000; Chen, Gully & Eden,

2001; Schwoerer, May, Hollensbe & Mencl, 2005). They hypothesize that, given enough exposure to successes in multiple specific domains, a person can develop a *broader concept* of self-efficacy that is applicable in a more global sense to a wide variety of behavioral outcomes.

Numerous studies have been published in recent years that observe the predictive ability of SSE and GSE in relation to performance. In large part, the evidence for the SSE-performance link has been surprisingly consistent across a number of different domains. A meta-analysis by Stajkovic and Luthans (1998) observed a weighted average correlation of .38 between SSE and work performance measures; a separate meta-analysis by Moritz et al. (2000) also noted an average correlation of .38, this time correlating SSE and performance in sport. Various other empirical studies have observed SSE-performance correlations of .37 (Mathieu, Martineau & Tannenbaum, 1993), .38 (Phillips & Gully, 1997), and .34 (Lane & Lane, 2001).

Compared to the previous findings, correlations between GSE and performance have been significantly less definitive. In one of the few meta-analyses available for this particular relationship, Judge and Bono (2001) reported an average corrected correlation of .23 between GSE and job performance, substantially smaller than the SSE-performance link described previously. Still other singular studies have failed to even find a single significant correlation. For example, Chen et al. (2001) found no relationship between students' GSE prior to taking an exam and subsequent performance (though measures of GSE after receiving exam scores did reveal significant correlations, reiterating the importance of performance accomplishments on self-efficacy development, Bandura, 1977); a similar finding was also reported by Chen et al. (2000).

Based on the evidence reviewed here, a measure of SSE seems the more appropriate approach for the purposes of the current study. Not only has it been shown to be a more reliable and potent predictor of performance, it is also relevant to the specificity of the remaining model variables and thus maintains a meaningful methodological consistency (Bandura, 1986; Stajkovic & Luthans, 1988; Chen, Gully and Eden, 2001). The SSE domain of interest for the present research, mechanical self-efficacy, will herein be defined as an individual's judgments of his/her capability to learn about, perceive and understand the operation of common physical principles and mechanical elements. Thus, the level of *mechanical* interest, knowledge and experience possessed by an individual should be more closely related to *mechanical* self-efficacy, which should ultimately be more predictive of performance on a *mechanical* comprehension test. Therefore, it is hypothesized that:

> Hypothesis 2: The relationship between mechanical interests, knowledge and experiences and performance on a test of mechanical comprehension will be mediated by mechanical self-efficacy.

Gender role identification. Broadly categorized, gender roles refer to socially constructed norms that describe the "psychological characteristics that equip [a person] for the tasks that their sex typically performs" (Wood & Eagly, 2002, p. 701). Gender roles have traditionally been depicted along a single, bipolar dimension ranging from masculinity on one end and femininity on the other; however, researchers have long since broadened this perspective to allow for the possibility that an individual may possess traits typical of both sexes (Constantinople, 1973). In her seminal article, Bem (1974)

presented the Bem Sex Role Inventory (BSRI), which contained two separate subscales for assessing both masculinity *and* femininity within an individual, thus suggesting that a person may possess the psychological characteristics of both males and females. In practice however, the BSRI still places gender roles on a single continuum, categorizing individuals as feminine, near-feminine, androgynous, near-masculine, and masculine.

Others have attempted to move even further from the typical dichotomous continuum. Spence, Helmreich and Stapp (1975) developed a measure similar to the BSRI, but used median splits on the separate masculinity and femininity scales in an attempt to develop a more two-dimensional model. This technique resulted in four unique gender roles: masculine (high masculine-low feminine), feminine (low masculine-high feminine), androgynous (high masculine-high feminine), and undifferentiated (low masculine-low feminine). Interestingly though, when the distribution of subjects based on this new two-dimensional taxonomy are compared with the distribution that would be expected using the five gender roles generated from the BSRI, a largely similar grouping of individuals emerged (see Table 2).

The data presented in Table 2 reveal a number of interesting conclusions. First, although it may be theoretically beneficial to conceptualize gender role identification as two dimensional with masculinity and femininity lying on different continuums (Constantinople, 1973), it appears that a unidimensional model is a reasonably accurate substitute in practice. As can be observed in Table 2, approximately 94% of Spence et al.'s (1975) masculine males and 90% of feminine females fall into Bem's (1974) near-masculine/masculine and near-feminine/feminine categories, respectively. Similarly, approximately 80% of feminine males and 97% of masculine females sorted into the

23

Table 2

Percentage of Subjects in Bem's (1974) Five Gender Role Categories for each of the Four Gender Roles Defined by Spence, Helmreich and Stapp (1975)

Spence et al.'s Categories	Bem's Categories				
	Feminine	Near-Feminine	Androgynous	Near-Masculine	Masculine
Males					
Masculine (64)	0.0	0.0	6.3	15.6	78.1
Feminine (30)	46.7	33.3	20.0	0.0	0.0
Androgynous (68)	5.9	1.5	45.6	25.0	22.1
Undifferentiated (72)	8.3	6.9	37.5	16.7	30.6
Females					
Masculine (30)	0.0	0.0	3.3	23.3	73.3
Feminine (104)	72.1	18.3	9.6	0.0	0.0
Androgynous (80)	13.8	10.0	47.5	18.7	10.0
Undifferentiated (56)	17.9	19.6	33.9	17.9	10.7

Note. Numbers in parentheses indicate the number of subjects within each gender role category in the Spence et al. study.

Table reproduced from Spence, J. T., Helmreich, R., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality & Social Psychology*, 32(1), 29-39.

expected gender role categories proposed by Bem (1974) as well.

Second, it is clear that there is a discrepancy in terms of how to categorize individuals when they report either high (androgynous) or low (undifferentiated) identification with *both* gender roles. As Bem (1974) operationalizes the term, an individual is considered androgynous if the standardized difference score between the masculinity and femininity subscales is near zero; thus, the majority of people identified as undifferentiated *or* androgynous by Spence et al. (1975) should also be categorized as androgynous under Bem's (1974) method. However, as Table 2 indicates, in no case do more than 50% of undifferentiated or androgynous individuals qualify as androgynous by Bem's (1974) standards. It appears then, that either the median split technique utilized by

Spence et al. (1975) is not able to categorize individuals with ambiguous gender roles reliably or the BSRI's requirements for establishing androgyny are not sufficient.

Building on the previous point, when subjects were measured using the two-dimensional model and no dominant gender role emerged (i.e. the undifferentiated and androgynous groups), the *actual* gender of the subject was a fairly good predictor of gender role as conceptualized in the unidimensional case. For example, close to 47% of males who grouped into Spence et al's (1975) undifferentiated category were scored as near-masculine or masculine according to Bem's (1974) measure. In fact, in only one of the four cases (androgynous females) did this "actual gender" to "gender role" pattern not hold consistent. Though these findings do not necessarily suggest that using gender as a proxy for gender role is always appropriate when gender role identification is ambiguous, they do suggest that actual gender is a reasonably accurate predictor of one's "gender-typical" characteristics. This holds significant meaning when gender role is related to behavioral outcomes—especially in terms of performance on a traditionally gender-advantaged task such as a test of mechanical comprehension (Antil & Cunningham, 1982; see section below on mechanical test performance).

Whether it is conceptualized as one- or two-dimensional, there is little argument that gender role identification has a major influence on people's beliefs and attitudes and, consequently, their behaviors and interests (Bem, 1974; Maccoby & Jacklin, 1974; Spence, Helmreich & Stapp, 1975; Wood & Eagly, 2002). Eccles (1987) summarizes this point nicely, stating:

> "Gender roles mandate different primary activities for men and women. If success in
> one's gender role is a central component of one's identity, then activities that fulfill this
> role should have high value and activities that hamper efforts at successfully fulfilling
> one's gender role should have lower subjective value." (p. 152)

The reasoning above makes clear two important considerations for the current study. First, the degree to which gender roles influence an individual will vary as a function of the strength and certainty of his/her gender role identification (Nash, 1979). Thus, the ability level of a strong masculine/weak feminine identified individual and a strong feminine/weak masculine identified individual should be significantly differentiated, whereas a less pronounced difference would be expected from individuals with ambiguous gender role identifications.

To clarify this point, consider the relationship between mechanical comprehension and gender role identification depicted in Figure 3. As mechanical comprehension has been shown to be a typically masculine quality (see below, Spence et al., 1975), it follows that the more strongly masculine-identified and less strongly feminine-identified an individual is the more likely they are to possess higher "levels" of mechanical comprehension. Alternatively, the more strongly feminine- and less strongly masculine-identified one is, the lower their level of the construct. Both of these statements are represented by the data points labeled "Masculine" and "Feminine" in Figure 3. Notice also that the level of mechanical comprehension of persons without a clearly identified gender role (the "Androgynous" and "Undifferentiated" data points) is relatively equal. In relation to this study, then, gender role identification will be treated as a two-dimensional, continuous variable in which the most extreme differences will be found between strongly and wholly gender role identified subjects.

Second, as Eccles (1987) indicates, the activities a person pursues should be significantly correlated with their most strongly identified gender role. This should also necessitate that the interests, knowledge and experiences that an individual accumulates

26
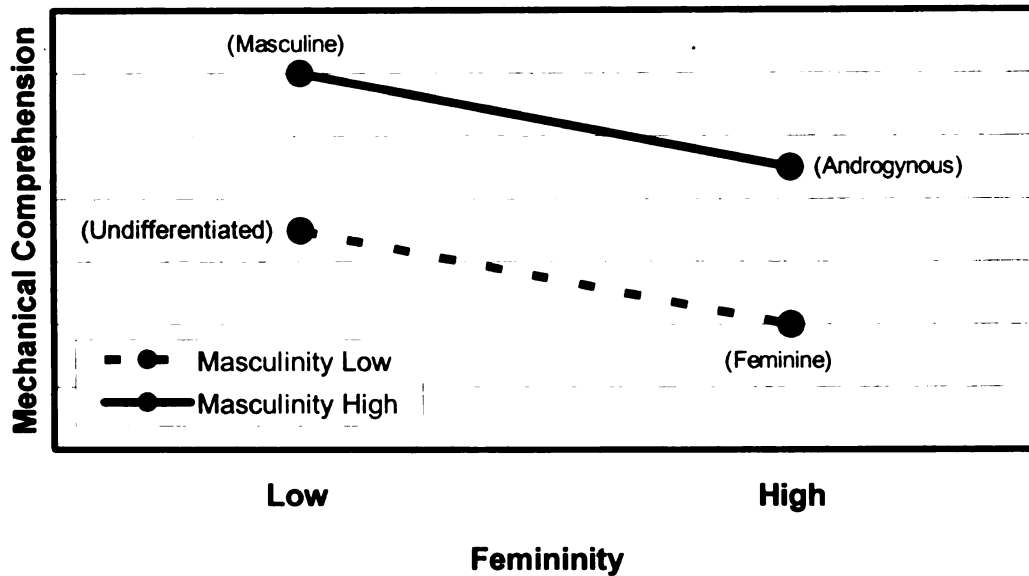
Figure 3. Example of the hypothesized effects of gender role identification on an individual's level of mechanical comprehension.

as a result of his/her primary activities should be consistent with their gender role. For example, one might expect that an individual, regardless of actual gender, who identifies with the psychological characteristics typical of femininity (i.e. affection, compassion, sympathy) would pursue activities that are traditionally considered feminine (i.e. cooking, caring for children), and to thus acquire certain interests, knowledge and experiences that a masculine-identified individual would most likely not (Bem, 1974; Wood & Eagly, 2002).

As it pertains to the present study, Spence et al. (1975) presented a number of descriptive statements to groups of college students and asked them to rate them as either male-valued, female-valued, or sex specific (meaning that the "ideal" male or female would wholly possess/perform the trait/activity, while the other gender would not). Of the sex specific items observed by the survey, mechanical aptitude emerged as one of the characteristics most typical of the ideal male and the masculine gender role. Therefore, it

is expected that participating in activities related to mechanical aptitude will be highly correlated with the masculine gender role, such that:

Hypothesis 3a: Individuals who identify more with a feminine gender role will have less interest in, less general knowledge of, and fewer experiences with mechanically-related subject matter.

Hypothesis 3b: Individuals who identify more with a masculine gender role will have more interest in, greater general knowledge of, and more experiences with mechanically-related subject matter.

Endorsement of gender stereotypes. Simply stated, a stereotype is a broad representation of the physical, psychological and behavioral similarities of a large group of people (Stangor, 2000). Stereotypes about gender, then, can be defined as "socially shared beliefs about the characteristics or attributes of men and women in general" (Cleveland, Stockdale & Murphy, 2000, p. 43). For example, Williams and Bennett (1975) found that 99% of males (n = 50) and 100% of females (n = 50) identified the adjective "emotional" as a typical characteristic of women; similarly, 98% of the male and 98% of the female sample labeled "forceful" a male-typical trait. Once a stereotype about a particular set of people has been established, its associated characteristics are often taken as true of any *individual* affiliated with that group (Cleveland et al., 2000, Stangor, 2000). Thus, based only on the gender stereotype, any female would be considered emotional and any male would be considered forceful. Such sweeping generalizations may result in a number of outcomes in the larger societal picture, some

more negative (i.e. prejudice, discrimination) than others (Devine, 1989; Lepore & Brown, 2000).

In focusing more specifically on gender stereotypes, researchers have captured the phenomena at two distinct levels of operation—*sex-role stereotypes* and *sex-trait stereotypes* (Williams & Best, 1990). Sex-role stereotypes refer to "beliefs about the appropriateness of various roles and activities for men and women" (Cleveland et al., 2000, p. 43), and sex-trait stereotypes refer to "beliefs that [specific] psychological and behavioral characteristics describe the majority of men to a greater or lesser degree than the majority of women" (p. 44). Based on these more explicit treatments, it is appropriate at this point to make an important distinction between sex-role stereotypes and the previously discussed gender role identification. Although the two seem very similar, a major difference exists in how they are conceptualized, which in turn has significant influence on their meaningfulness to the current research question.

Sex-role stereotypes are defined as collective judgments of the *appropriateness* of the skills, activities, roles, etc. that are associated with a particular gender. On the other hand, gender role identification refers to the degree to which an individual believes the skills, activities, roles, etc. *that he/she possesses* are representative of the tasks and behaviors their sex typically performs (Bem, 1974; Wood & Eagly, 2002). Perhaps at the risk of oversimplifying the issue, a sex-role stereotype can be perceived as a cultural directive depicting whether it is appropriate for a male to possess feminine-typical characteristics and a female to possess masculine-typical characteristics; conversely, gender role identification can be interpreted as a personal endorsement of whether one actually has masculine or feminine characteristics (irrespective of sex or societal

acceptance). This is important in terms of the present research because although sex-role stereotypes can have self-limiting consequences (Cleveland et al., 2000), such effects depend greatly on the degree to which an individual actively "buys into" the stereotype (Eccles, 1987; Nosek, Banaji & Greenwald, 2002) or is aware that a stereotype is applicable in a given situation (Spencer, Steele & Quinn, 1999). Based on the preceding, it is reasonable to suspect that gender role identification (thought to be a more stable and enduring individual quality, Constantinople, 1973; Bem, 1974) would be a more valid and reliable predictor of the individual differences related to sex roles. For the purposes of this study then, gender stereotypes will only be considered from the perspective of sex-trait stereotypes.

Well over fifty years of research has attempted to document the traits and characteristics that are stereotypically associated with men and women, and a large number of such factors have been identified that are surprisingly consistent over time and culture (Williams & Best, 1990). Although these classifications are interesting and useful in their own right, the current experiment is more concerned with the individual outcomes associated with endorsing gender stereotypical beliefs. To this end, Cleveland et al. (2000) pose an interesting question—might the stereotypical traits of a given sex be negatively related to the attainment of certain desirable outcomes? Given the wide range of traits that stereotypically distinguish men from women, researchers have suggested that there are any number of "male" or "female" behavioral and psychological characteristics that can explain why various performance, ability, etc. differences exist (Williams & Best, 1990). For example, achievement motivation has been found to be a reliable predictor of job performance across a number of studies (Piedmont & Weinstein,

1994; Vinchur, Schippmann, Switzer & Roth, 1998; Stewart, 1999; Van den Berg & Feij, 2003). However, Stein and Bailey (1973) demonstrated that men generally exhibit significantly higher levels of achievement motivation than women over the course of the natural lifespan. Though there may be a number of reasons as to why this difference exists, Cleveland et al. (2000) suggest that one possible explanation can be derived by observing the stereotypical sex traits associated with females in relation to achievement motivation; across nearly all situations, feminine characteristics (such as non-assertiveness, avoidance of competition, and dependency) are in direct conflict with the necessary requirements of achievement motivation (e.g. striving for success in all situations, desire to be "better" than others, etc.). This line of reasoning suggests that the observed disparity in performance between males and females on a particular task could be attributed to differences in achievement motivation, which is negatively influenced by one's adherence to traditionally feminine characteristics.

This notion of "compatibility" between feminine/masculine characteristics and attitudes, behaviors, and cognitions as a driver for the observed gender differences across various areas of psychological functioning has been labeled *role congruity theory* by Eagly and Karau (2002). However, there are two subtle, yet important, components to the arguments for role congruity theory that are necessary to understand in order to adequately discuss the present study's model of gender differences in mechanical comprehension. First, gender role identification is not proposed to directly impact performance; instead, gender role identification affects *characteristics of the performance context* (e.g., achievement motivation), which then subsequently affect performance outcomes (cf. Cleveland et al., 2000). In regards to this study's model, the intervening

31

variable proposed to be affected is mechanical self-efficacy. Based on the rationales developed for Hypothesis 2 and Hypotheses 3a/3b, individuals with more masculine qualities are predicted to possess traits and levels of mechanical interest, knowledge and experience that predispose them to be more efficacious when it comes to understanding mechanically-related concepts than more feminine individuals. Further supporting this reasoning is the finding reported by Spence et al. (1975) that mechanical aptitude is generally perceived as a stereotypically male valued characteristic. Owing to its "male heritage" then, masculine-identified individuals should perceive that they are better equipped (i.e., more role congruous) to deal with mechanically-related activities and thus report more confidence in their ability to identify and complete mechanical tasks.

The second, less immediately obvious component of this argument is actually an important caveat to role congruity theory: even in situations in which the characteristics of a particular gender role are not conducive to a certain favorable outcome (e.g., high femininity to high achievement motivation), individuals who identify with that gender role *will not always* be at a marked disadvantage to their counterparts. For example, a significant amount of research has been directed at examining the differential outcomes associated with male versus female leaders (cf. Brown, 1979; Eagly, Johannesen-Schmidt, van Engen & Marloes, 2003). Similar to the point made in the preceding paragraph, the assumption often examined in the literature is that the qualities which characterize the feminine gender role are incongruous with those required to be successful as a leader, and thus females have a much more difficult time becoming or sustaining leadership positions (Eagly & Karau, 2002). Although evidence has been found in support of this proposition, meta-analyses have shown that a number of contextual and individual difference

characteristics moderate this relationship such that feminine leaders are not always passed over or perceived as inferior to masculine leaders (e.g., van Engen & Willemsen, 2004).

What this suggests, then, is that the relationship between gender role identification and performance-related variables can be moderated by a variety of characteristics at differing levels of analysis. In maintaining the individual-level model proposed in Figure 2, the moderating variable of interest examined by the current study is gender stereotype endorsement. The rationale for identifying stereotype endorsement as a moderator is based on the extensive stereotype threat literature (e.g., Steele & Aronson, 1995; Schmader, 2002; Schmader, Johns & Barquissau, 2004), in which the presence and salience of a self-important, negative stereotype results in a member of the stereotyped group underachieving.

In the simplest stereotype threat study, individuals are either told or not told that a particular test has historically shown significant gender, racial, cultural or other group differences (i.e. a stereotype is made salient). Results from such studies typically find that in situations where the "threat" or bias is activated, the performance differences between the stereotyped-group and the neutral-group are much greater than when the manipulation is not present. Thus the model of stereotype threat assumes that people enter a situation with differing degrees of interest, knowledge and experience relevant to the test that might normally predict test performance. But, when the presence of a test bias is known, this relationship is negatively affected for the predicted group by an intervening variable. This situation is quite similar to what is hypothesized by the current

study, and thus it is theoretically justifiable to predict stereotype endorsement as a moderating variable.

However, unlike stereotype threat research, the present study deals with a person's *endorsement* of gender stereotypes rather than their *awareness* of the stereotype. In order to activate a stereotype threat effect, Steele and Aronson (1995) state that an individual must "face the threat of confirming or being judged by a negative societal stereotype—a suspicion—about their group's intellectual ability and competence (p. 797)." In this sense, the stereotype acts only as an arbitrary frame of comparison. Stereotype threat is not concerned with whether an individual *believes* the difference between their subgroup and others is true; instead, it focuses on the processes through which an individual actively avoids fulfilling the stereotype.

This distinction is often presented in the prejudice/discrimination literature as the difference between stereotype endorsement and stereotype knowledge (Devine, 1989; Lepore & Brown, 2000). Devine (1989) argues that these two constructs are "conceptually distinct cognitive structures...[and thus represent] only potentially overlapping subsets of information" (p. 5). Furthermore, to the extent that both stereotype knowledge and endorsement are conceptually distinct, they likely have different implications for the expression and development of related beliefs about the abilities of oneself and others. In relation to the current study then, the moderating effect of one's belief in a stereotype (i.e., endorsement) is predicted to affect performance by influencing levels of mechanical self-efficacy, or one's belief in their mechanical proficiency. This claim simultaneously maintains the well-established SSE-to-

performance relationship developed earlier in Hypothesis 2, while still accounting for the individual influence of gender stereotype endorsement.

Thus, as the model being tested by the current study follows a basic mediation model in which mechanical self-efficacy is the link between mechanical interest, knowledge, and experience and performance on a mechanical comprehension test, it is hypothesized that endorsement of gender stereotypes will affect performance by moderating the relationship between one's gender role identification and his/her reported level of mechanical self-efficacy. As was discussed earlier, gender role identification is closely related to the sex-role stereotype component of gender stereotypes—thus, a significant interactive relationship is expected to emerge between these two constructs (Figure 4). Though no evidence was found in the literature to support this exact claim, a significant Sex (male or female) by Implicit Stereotype Endorsement (male = math, female ≠ math) interaction effect on math test performance was reported by Nosek et al. (2002), indicating that the interaction predicted by the present study is likely.

In summary, it is hypothesized that endorsement of gender stereotypes will moderate the relationship between gender role identification and mechanical self-efficacy, such that:

> Hypothesis 4: A significant interaction between gender role identification and gender stereotype endorsement will emerge such that feminine individuals who more strongly endorse gender stereotypes will report lower levels of mechanical self-efficacy than those who do not endorse gender stereotypes. The interaction will not be observed for masculine identified individuals.
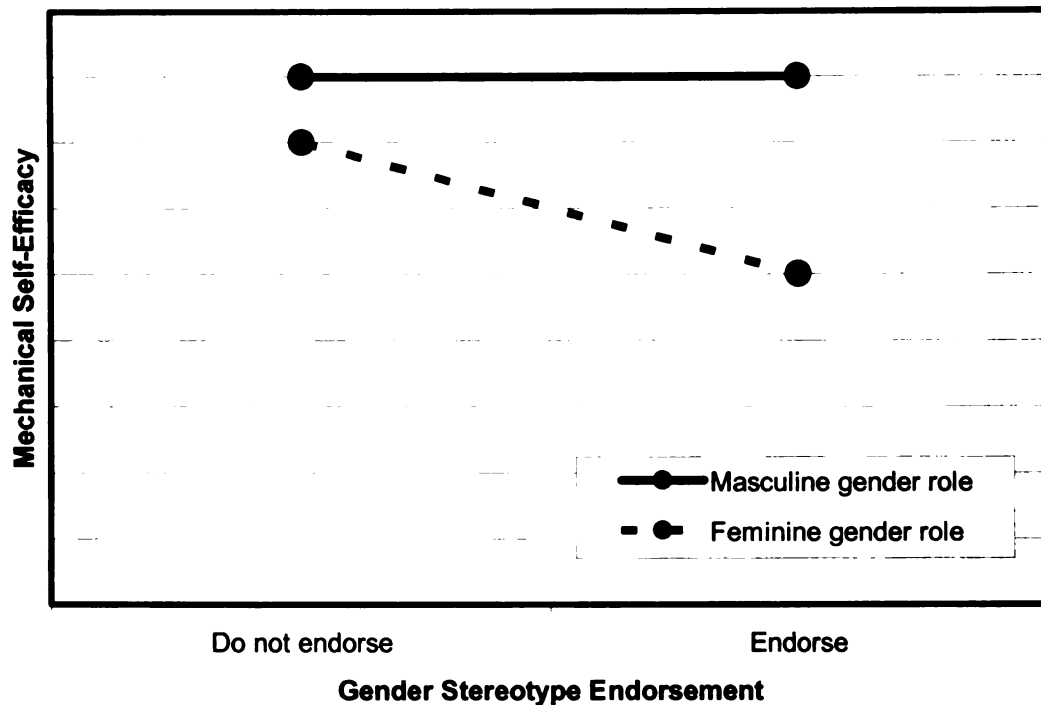
Figure 4. Predicted interaction effect of gender role identification and gender stereotype endorsement on reported levels of mechanical self-efficacy.

Performance on test of mechanical comprehension. The final hypotheses made by the current study concern actually predicting performance on the mechanical comprehension test. Based on Hypothesis 2, it is predicted that performance will be directly related to mechanical self-efficacy, a finding that has been strongly supported by the literature (Stajkovic & Luthans, 1988; Mathieu et al., 1993; Phillips & Gully, 1997; Moritz et al., 2000; Lane & Lane, 2001); but as the previous discussions have explicated, many factors are predicted to influence mechanical self-efficacy in a number of ways and combinations, the effects of which should ultimately be seen in overall test performance. However, none of the studies reviewed included mechanical self-efficacy as a predictor of mechanical performance, instead opting to look at the singular, causal effects of the gender-related variables on performance. Thus, in lieu of a more direct correlate of this

study's proposed model, understanding how these various predictors relate directly to mechanical test performance offers the best means for justifying the performance predictions hypothesized by the present research.

In one of the earliest studies attempting to capture performance differences based on gender and gender roles, Milton (1957) compared the scores of 63 males and 66 females with scores on a test of general "problem-solving" skills that had shown significant gender differences previously. Milton succeeded in reproducing the significant overall effect for gender (in which males performed better on the test than females), but of even greater interest, he also found that sex-role identification accounted for up to 20% of the variance in the male sample's test scores and 7% of the variance in the female sample. As further evidence of the importance of sex roles, when scores on the gender-role identification measure were introduced as a covariate in the regression equation, the previously significant sex difference in problem-solving scores dropped to non-significance. Though it is unclear whether the problem-solving skill Milton measured is specifically related to mechanical comprehension, these findings nevertheless lend significant support to the influence of gender role identification on certain cognitive tasks.

Of a more direct significance to mechanical comprehension and the current study, Nash (1975) conducted an experiment with sixth- and ninth-grade boys and girls to observe the relationships between gender, gender role preference and spatial ability. Subjects in the study were asked to complete an open-ended questionnaire regarding gender role preference that asked them to explain whether they thought it better to be a male or female and whether they would rather be a male or female. Spatial ability was

assessed with scores on the Differential Aptitudes Space Relations Test. Nash found the consistently reported male advantage in spatial ability that had been noted by other researchers (Maccoby & Jacklin, 1974); however, an interesting pattern emerged when gender was crossed with gender role preference (Table 3). Compared by their similar age groups, sixth- and ninth-grade boys who preferred to be boys substantially outperformed sixth- and ninth-grade girls who preferred to be girls on the spatial ability test. However, in both age groups, no significant difference in spatial test scores emerged between girls who indicated they would prefer to be boys and boys who preferred to be boys. This finding seems to suggest that the interaction between gender and gender role identification may be a much better indicator than gender alone in predicting performance on tests of specific cognitive ability, even in cases where marked gender differences have been observed. Nash's study holds importance for the present research question, as mechanical comprehension has long been known to load on the spatial visualization factor (Guildford, 1947). Therefore, it is reasonable to expect similar findings to emerge when mechanical comprehension is used as the criterion measure.

Antill and Cunningham (1982) present perhaps the most direct—though

Table 3

Mean Spatial Reasoning Performance (SD) on the Differential Aptitudes Space Relations Test as a Function of Gender and Gender Preference

6th-Grade boys who prefer to be boys (n = 32)	9th-Grade boys who prefer to be boys (n = 25)	6th-Grade boys who prefer to be girls (n = 4)	9th-Grade boys who prefer to be girls (n = 0)	6th-Grade girls who prefer to be girls (n = 32)	9th-Grade girls who prefer to be girls (n = 41)	6th-Grade girls who prefer to be boys (n = 23)	9th-Grade girls who prefer to be boys (n = 7)
22.96 (15.21)	42.26 (19.87)	6.75 (12.20)	--	12.91 (13.07)	29.04 (18.59)	21.75 (10.59)	44.19 (17.36)

Table reproduced from Nash, S. C. (1975). The relationship among sex-role stereotyping, sex-role preference, and the sex difference in spatial visualization. *Sex Roles, 1*(1), 15-32.

incomplete—examination of the effects of gender and gender role identification on

mechanical comprehension test performance. Here, the researchers administered five

different gender-role identification measures and the ACER Mechanical Reasoning Test

to 237 participants (56% female). The sample was then separated into the four gender

role groups specified by Spence et al. (1975)—masculine, feminine, androgynous, and

undifferentiated—using a median split technique to produce a 2 (gender) x 4 (gender role)

x 5 (measure) matrix. The observed cell means from this experiment are reproduced in

the last four columns of Table 4.

Table 4

Mechanical-Reasoning Means Based on Median Split Method Categories

Tests	Category						
	M + A	U + F	t	M	A	U	F
Males							
BSRI	17.1 (68)	15.9 (36)	1.46*	17.0 (46)	17.2 (22)	16.1 (27)	15.2 (9)
PAQ	17.1 (65)	15.9 (39)	1.43*	18.1 (32)	16.1 (33)	16.0 (33)	15.3 (6)
ANDRO	16.9 (79)	15.8 (25)	1.27*	17.1 (46)	16.8 (33)	16.4 (14)	15.0 (11)
CPI	16.8 (78)	16.2 (26)	.72	17.2 (49)	16.2 (29)	17.1 (18)	13.9 (8)
Comrey	17.2 (79)	14.9 (25)	2.52***	17.1 (64)	17.7 (15)	13.8 (8)	15.5 (17)
Females							
BSRI	11.9 (51)	11.7 (82)	.39	12.5 (19)	11.6 (32)	11.8 (21)	11.7 (61)
PAQ	11.7 (51)	11.9 (82)	.35	13.9 (14)	10.8 (37)	12.2 (31)	11.7 (51)
ANDRO	12.4 (51)	11.5 (82)	1.50*	12.0 (20)	12.7 (31)	12.6 (12)	11.3 (70)
CPI	12.6 (49)	11.3 (84)	2.10**	13.8 (12)	12.2 (37)	10.8 (30)	11.6 (54)
Comrey	12.6 (39)	11.5 (94)	1.81**	12.5 (24)	12.7 (15)	11.6 (11)	11.4 (83)

$* p < .10. ** p < .05 *** p < .01$, one-tailed for the directional hypothesis.

Note. Full names of tests are: BSRI = Bem Sex-Role Inventory; PAQ = Personal Attributes
Questionnaire; ANDRO = androgyny scale for the Personality Research Form; CPI =
California Psychological Inventory; Comrey = Comrey Personality Scale. Full names of
categories are: M = masculine; A = Androgynous; U = Undifferentiated; F = Feminine.
Numbers in parentheses indicate the number of subjects on which the mean next to it is
based.
The t column refers to the results of the t-test between the M + A and U + F groups in that
row.

Table reproduced from Antill, J. K., & Cunningham, J. D. (1982). Sex differences in performance
on ability tests as a function of masculinity, femininity, and androgyny. *Journal of Personality &
Social Psychology, 42*(4), 718-728.

Unfortunately, these authors do not present an explicit test of the gender by gender role interaction in their results, instead focusing on a broader examination of the main effects of gender and gender role identification. That being said, significant effects were found for both gender and gender role identification. As expected, males outperformed females by a significant margin ($F(1,235) = 100.9$, $p < .001$); additionally, the measures of gender role identification also accounted for significant variance in mechanical test performance, even when controlling for gender ($F(5,230 = 2.81$, $p < .025$). An examination of the cell means across gender revealed that masculine individuals outperformed both feminine and undifferentiated individuals in 19 of 20 cases; furthermore, in 15 of 20 instances androgynous individuals also scored higher than these latter two groups. Also of note, the masculine group outperformed the androgynous group in 6 of 10 comparisons and the undifferentiated group scored better than the feminine group in 8 of 10 cases.

Antill and Cunningham (1982) also attempted to examine mechanical test performance within masculine identified individuals (first three columns of values in Table 4). The authors began by aggregating the four gender roles (masculine, feminine, androgynous and feminine) into high-masculine and low-masculine groups for both males and females separately using a median split technique. The high-masculine group was constructed by combining the masculine (characterized by Spence et al., 1975, as high masculine/low feminine identified) and androgynous (high masculine/high feminine) participants together, whereas low-masculine merged feminine (low masculine/high feminine) and undifferentiated (low masculine/low feminine) persons into a single group. Dividing the sample in this way resulted in the formation of high-masculine/low-

40

masculine groups composed only of males and high-masculine/low-masculine composed only of females (see Table 4).

t-tests were then performed within gender across the masculine differentiated groups. As can be seen from their results, the analyses were somewhat inconclusive and dependent upon the gender role measure implemented. Although the observed mean scores of the more masculine group were higher than those of the less masculine group in all but one case, the differences were only significant (at the .05 level) in one comparison for males and in two comparisons for females. Regrettably, although their dataset allowed for it, a similar analysis was not conducted for high- (feminine + androgynous) versus low-feminine (undifferentiated + masculine) groups. Had such an analysis been conducted, the presence of a consistent significant difference between high- and low-feminine males versus high- and low-feminine females would have suggested a possible gender by gender role interaction. As presented though, the observed differences appeared to show that higher masculine-identified individuals performed better on the mechanical comprehension test than lower masculine-identified individuals irrespective of gender, though this conclusion is somewhat suspect given the lack of reliable findings across measures of gender role.

Taken together, these findings reveal important insights into the effect of gender role identification and mechanical comprehension. First, as found in previous studies, identification with the masculine gender role did appear to correlate with the highest performance on male-advantaged tests of specific cognitive ability. Second, the Antill and Cunningham (1982) findings also suggest that femininity may have a detrimental effect on mechanical comprehension. In the comparisons where masculinity was a

constant (i.e. masculine versus androgynous, undifferentiated versus feminine), the group that was lowest in femininity was the better performer. Finally, although Antill and Cunningham did not explicitly test for a gender by gender role interaction, their findings do hint that such an interaction may exist. For example, while no analysis was presented to test for significant mean differences across cell means, an examination of their absolute values (excluding the androgynous and undifferentiated categories) shows that masculine males were always the top performers, followed by feminine males, masculine females and finally feminine females[1].

While the studies summarized above represent some of the few that have looked at the effects of gender and gender role identification on specific ability testing, even less research has considered the role of gender stereotype endorsement on such test performance. What research is available generally indicates that stereotype endorsement is not a significant main effect predictor of performance. Only in cases where stereotype endorsement acts as part of an interaction effect does it emerge as a significant predictor of performance. For example, Schmader et al. (2004) reported that although stereotype endorsement was not a significant predictor of a female's performance on a mathematics test by itself, the stereotype endorsement by stereotype threat interaction was highly predictive ($\beta = -.30, p < .01$). Similarly, Nosek et al. (2002) found that the interaction of implicit stereotypes pertaining to gender and math with gender was a significant predictor of math performance ($\beta = -.31, p < .01$), though, again, stereotype endorsement alone was not significant ($\beta = .14, p = ns$).

---

[1] Because males were the high scorers in all instances, the earlier arguments that proposed a biologically-driven (such as with spatial ability, Harris, 1978; 1981) or socializing (Eccles, 1987) influence predisposing men to be better at tasks of mechanical comprehension than women still appear to be strong factors.

Based on the preceding, it is predicted that a significant three-way interaction between gender, gender role identification and gender stereotype endorsement on performance will emerge as indicated in Figure 5. Recalling the Spence et al. (1975)

Performance on mechanical comprehension test as a function of gender and stereotype endorsement with feminine gender role identification

Performance on mechanical comprehension test as a function of gender and stereotype endorsement with masculine gender role identification
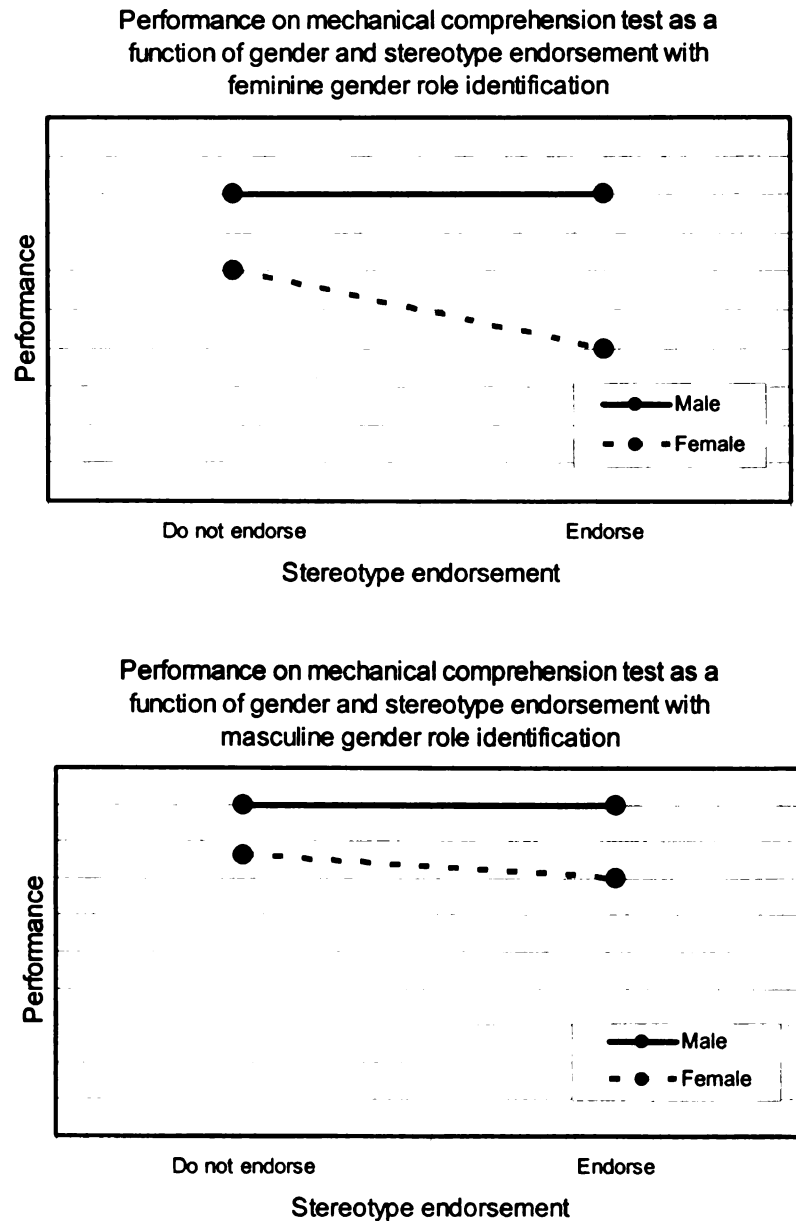
Figure 5. Predicted three-way interaction effect of gender, gender role identification and gender stereotype endorsement on test performance.

finding that mechanical comprehension is a sex trait stereotypically associated with males and that gender stereotypes pertain to *gender* and not *gender roles*, it should follow that males will be unaffected by gender stereotype endorsement. However, this should not hold true in the opposite case; endorsement of gender stereotypes should differentially affect females such that endorsement results in lower test performance than no endorsement. In addition, within gender, masculine gender role identification should result in greater test performance than feminine gender role identification (Antill & Cunningham, 1982). In summary then, it is predicted that:

Hypothesis 5: There will be a three-way interaction between gender, gender role identification, and gender stereotype endorsement such that women who are feminine (or masculine) identified and endorse gender stereotypes will score lower than feminine (or masculine) identified women who do not endorse gender stereotypes; this pattern will not be evidenced for males (see Figure 5).

As gender stereotype endorsement is predicted to differentially interact with gender, the remaining hypotheses concern the interaction between gender and gender role identification. To examine this relationship, an attempt will be made to explicitly test the conclusions implied in the research by Antill and Cunningham (1982), which stated that masculine males achieved the highest scores on a test of mechanical ability, followed by feminine males, masculine females and feminine females.

Hypothesis 6a: Males who identify with a masculine gender role will score the highest on the test of mechanical comprehension.

<u>Hypothesis 6b</u>: Males who identify with a feminine gender role will score lower on the test of mechanical comprehension than masculine males, but higher than masculine females and feminine females.

<u>Hypothesis 6c</u>: Females who identify with a masculine gender role will score higher on the test of mechanical comprehension than feminine females, but lower than masculine males and feminine males.

<u>Hypothesis 6d</u>: Females who identify with a feminine gender role will score the lowest on the test of mechanical comprehension.

METHOD

Sample

258 participants from undergraduate introductory psychology courses at a large Midwestern university enrolled in the study as partial fulfillment of course requirements; Table 5 presents descriptive statistics for the sample. Involvement was voluntary, though all individuals who participated received class credit. Prior to testing, effort was made to ensure that the sample was approximately gender balanced in order to ensure that differences between- and within-gender could be adequately examined. As Table 5 depicts, only 38% (n = 99) of the total sample were male; this slight skew, though, was expected given the gender distribution of the available subject pool. However, a priori power analyses predicting a small effect size (d = .2) indicated that 84 to 102 subjects from each sex (168 to 204 total participants) would be adequate to find statistically significant results given this study's hypotheses thus the unequal cell sizes should not be problematic.

Measures

Demographics/Background. Standard demographic variables were obtained from participants in a brief online survey. Items for the demographics/background measure are presented in Appendix A.

Mechanical interests, knowledge and experiences. Mechanical interests, knowledge and experiences were assessed using an 18-item survey adapted from the Mechanical Experiences Background Questionnaire (MEBQ, Rechenberg, 2000). The MEBQ is a biodata measure specifically constructed to assess mechanical trainability. It

Table 5
Demographic Information of Study Sample (N = 258)

Gender	
Male	99
Female	159
Age (*n = 248*)	$M = 19.83$ (*SD* = 1.63)
Ethnicity (*n* = 255)	
American Indian/Alaskan Native	2
Asian	13
Black or African American	32
Native Hawaiian and Other Pacific Islander	1
White	199
American Indian/Alaskan Native and White	1
Asian and White	3
Black or African American and White	4

originally consisted of 93 items grouped into 13 dimensions derived from the job descriptions and job analyses of mechanically oriented jobs found in the Dictionary of Occupational Titles. However, as the thesis of the current study is concerned with the more cognitively-focused construct of mechanical comprehension rather than the more behaviorally-focused construct of mechanical trainability, many of the MEBQ's dimensions are not applicable (such as Following Directions and Coordination). Thus items were selected from the MEBQ that were judged to most closely conform to the three dimensions evaluated in the present study (interests, knowledge and experiences). The wording of items was then adapted such that a 5-point Likert-type scale ranging from "*Strongly Disagree*" to "*Strongly Agree*" could be used to answer each question. Items for the complete scale can be found in Appendix B.

Mechanical self-efficacy. Mechanical self-efficacy was assessed with an 8-item measure specifically constructed for the purposes of this research project. To develop the measures, item content was written to appropriately reflect the definition of mechanical comprehension presented earlier; namely, a person's ability to learn about, perceive and understand physical principles and mechanical elements in everyday situations. Participants were asked to indicate their confidence level regarding their ability to perform tasks related to mechanical comprehension on a continuous 7-point scale ranging from "*Not at all confident*" to "*Completely confident.*" The complete mechanical self-efficacy scale is shown in Appendix C.

Gender role identification. Gender role identification was assessed with the Bem Sex Role Inventory (BSRI, Bem, 1974). The BSRI asks individuals to indicate how accurately 60 personality characteristics (20 masculine, 20 feminine and 20 neutral items[2]) describe themselves on a 7-point scale ranging from "*Never true of me*" to "*Always true of me.*" Both a Masculinity and Femininity subscale can then be calculated by combining respondents' answers on the respective items. Bem (1974) reports the internal reliability for scores on the Masculinity items to be $\alpha = .86$, and $\alpha = .80$ for items measuring Femininity. In addition, test-retest reliabilities for both sets of scores were also found to be quite high (Masculinity $r = .90$; Femininity $r = .90$). Items for this measure are presented in Appendix D.

---

[2] Bem (1974) does not make explicit the manner in which the neutral items of the BSRI subscale relate to items from the masculinity and femininity subscale; given that they are posited to be neutral, the presumption is that they would not share strong relations. However, an examination of these items in Appendix D reveals that they likely they do not hold such a neutral valence. The items describe both "positive" (happy, friendly, helpful) and "negative" (moody, jealous, secretive) aspects of one's personality, which, depending upon somebody's preferred gender role, may load differentially on the masculine/feminine subscales. Owing to its ambiguous nature then, this subscale is not examined further in the present study.

Gender stereotype endorsement. Gender stereotype endorsement was assessed with a 10-item stereotype endorsement measure adapted from Blanton, Christie and Dye (2002) and Levy, Stroessner and Dweck (1998). From Blanton et al. (2002), participants are presented with a brief list of common gender stereotypes and asked to rate the degree to which they personally believe the stereotypes are based on true differences between males and females. Levy et al. (1998) proposed ratings be obtained on a 5-point scale ranging from *"Not at all true"* to *"Absolutely true."* In addition, Levy et al. suggest providing an example of a true stereotype to reduce a participant's tendency to answer in a completely socially desirable manner and report all stereotypes as false. Purposeful emphasis was placed on having individuals rate the degree to which "I believe..." a specific stereotype is true in order to ensure that the measure was one of stereotype endorsement and not stereotype knowledge (Devine, 1989; Lepore & Brown, 2000). In a further attempt to disentangle stereotype knowledge from the measure, the content of the items was specifically chosen to represent commonly recognized gender stereotypes, with the presumption being that the vast majority of respondents should already know about the purported gender difference and thus stereotype knowledge would be relatively equal across all participants. Items for the gender stereotype endorsement measure can be found in Appendix E.

Mechanical comprehension test performance. Mechanical comprehension was assessed with the most recent version of the Bennett Mechanical Comprehension Test Form S (BMCT, Bennett, 2006). The BMCT is a timed 30 minute, 68-item, multiple-choice test published by Harcourt Assessment and is widely considered to be one of the most frequently administered psychometric tests of mechanical ability (Muchinsky, 2004).

The test was specifically developed to "measure [one's] ability to perceive and understand the relationship of physical forces and mechanical elements in practical situations" (Bennett, 1969, p. 1). The BMCT contains items distributed across 18 categories of commonly encountered mechanical phenomena: Acoustics, Belt Drives, Center of Gravity, Centrifugal Force, Electricity, Gears, Gravity and Velocity, Heat, Hydraulics, Inertia, Levers, Optics, Planes and Slopes, Pulley Systems, Resolution of Forces, Shape and Volume, Structures, and Miscellaneous. Each question consists of a stem and a corresponding illustration, followed by three response options from which test takers are to choose the correct answer. Figure 6 provides an example of two questions taken from an earlier version of the test.

The BMCT manual for Form S (Bennett, 1969) reports published split-half
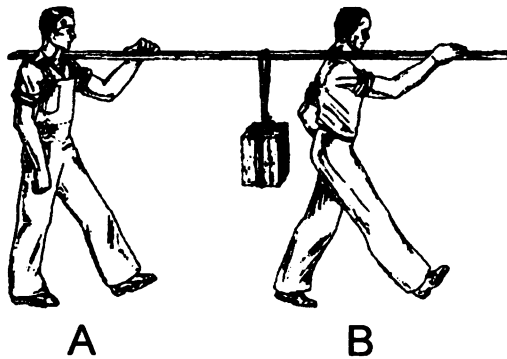
X

Which man carries more weight?
(If equal, mark C.)

A   B

A   B   C
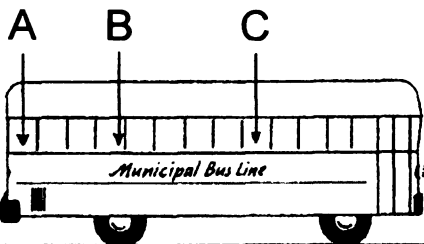
Y

Which letter shows the seat where a passenger will get the smoothest ride?

Source: Bennett, G. K. (1969). *Bennett mechanical comprehension test, Form S.* New York: The Psychological Corporation.

Figure 6. Example items from BMCT Form S

reliabilities corrected by the Spearman-Brown formula ranging from .81 to .93 across a variety of samples, with a median value of .86. To obtain an overall score of mechanical ability, each item is marked correct or incorrect and the total number of correct responses is simply summed; no subscales are computed from the BMCT.

Procedure

Study participants were recruited from the psychology department subject pool to take part in the study. In an effort to reduce the amount of time required of subjects, all surveys except for the BMCT were administered online. Prior to beginning the experiment, all subjects were required to read and agree with an informed consent form (Appendix F) that notified participants of the procedures of the experiment, the compensation they would receive, and their rights as participants. After providing consent, all participants completed the online surveys in the same order, first responding to the Demographics questionnaire, followed by the Mechanical Interests, Knowledge and Experience survey, Mechanical Self-Efficacy survey, the BSRI, and the Gender Stereotype Endorsement survey. Subjects were then automatically scheduled to come to a classroom one week later to take a paper-and-pencil version of the BMCT. The one week break between completion of the online surveys and completion of the mechanical comprehension test was deemed sufficient enough time to negate any implicit and inadvertent stereotype threat effects that may have been engaged when filling out either the Gender Role Identification or Gender Stereotype Endorsement measures (Steele & Aronson, 1995).

Participants were administered the BMCT in small groups of approximately 10 to 25 individuals. Pencils and scantrons were provided, and the experimenter kept track of

the 30 minute time limit on a small handheld egg timer. After all participants had entered

the classroom, test materials were distributed and the following directions provided:

> "You will be taking a 30 minute, 68-item test of mechanical ability – please bubble in all
> your answers directly on the scantron and leave no marks in the test booklets. When you
> are done with the test, return your scantrons, booklets and pencils to the experimenter. If
> you finish the test before the timer goes off, you may turn in your materials and leave."

During the test, respondents were notified when there were 15, 10 and 5 minutes

remaining to complete the test. Upon finishing and handing in their test materials,

participants were offered a debriefing sheet (Appendix G) that outlined the intended

purpose of the study and how the data gathered from the participants would be used.

RESULTS

Scale Analyses

Unless otherwise noted, all analyses were conducted in SPSS 13.0. Prior to hypothesis testing, an extensive examination of the measures used in the study was undertaken to evaluate the structure and functioning of the scale items. As many of the scales administered were either created or adapted specifically for the purposes of the present research, it was important to determine whether the instruments were operating correctly and as expected. In order to do so, the response values for all reverse coded items were first changed to ensure that all questions within a scale referenced the intended construct in the same direction. Following this, scale reliabilities were calculated to determine if deletion of any of the scale items could improve the overall alpha level of the measure. Finally, the measures were subjected to exploratory factor analysis to investigate whether the empirically-derived factor structure corresponded with the theoretically proposed use of the scales. Based on the interpretation of the factor structure, any changes regarding the use of the scales in subsequent data analyses were made accordingly. The results of these efforts are described in the sections below.

It should be noted that a complete scale analysis is not presented for the BSRI and BMCT. As both of these scales are proprietary instruments with a long history of normative and specific instructions regarding their use and interpretation (Bem, 1974; Bennett, 1969), it was deemed appropriate to use the measures "as is" and thus not conduct an intensive exploration of the scales. Therefore, only scale reliabilities were calculated for both, with no subsequent changes made to either measure.

Mechanical Interests, Knowledge, and Experiences. Items six, four and four in the mechanical interests, knowledge and experiences questionnaires, respectively, were negatively-keyed and thus the response values for these items were changed to coincide with the remaining survey questions. Reliability estimates were then calculated for each scale independently. Initially, the alpha levels for the scales were as follows: mechanical interest, $\alpha = .81$; mechanical knowledge, $\alpha = .73$; and mechanical experiences, $\alpha = .88$. Examination of the reliabilities at the item level, however, indicated that the removal of certain items could improve overall scale reliability. Removing items one and six from the mechanical interests scale resulted in an overall $\alpha = .86$, removing item four from the mechanical knowledge scale improved reliability to $\alpha = .74$, and removing item four from the mechanical experiences scale achieved an alpha of $\alpha = .89$.

It should be noted that within all three scales, removal of the reverse coded items always resulted in a small improvement in reliability, suggesting that respondents did not answer the negatively-keyed items in the expected direction. Even after recoding the negatively-keyed items in the correct direction, a subsequent factor analysis using the full scales revealed an independent factor completely identified by the reverse coded items. Based on these analyses, two possible conclusions seem plausible. First, it is possible that the negatively-keyed questions were poorly constructed, such that they either did not translate into a direct opposite appraisal of the trait or were tapping a different construct altogether. However, this seems fairly unlikely given that the three reverse coded items addressed substantially different content (see Appendix B) and yet still loaded heavily on a single factor.

The second and perhaps more likely explanation is that the questions were not carefully read and attended to by study participants. Many examples exist in the research literature demonstrating how careless responding to questions in a survey that contains both positively- and negatively-keyed items will often result in a separate factor completely composed of the reverse coded items (cf. Schmitt & Stuits, 1985). As Schmitt and Stuits (1985) define it, careless responding occurs when a respondent simply reads a few of the items in a given scale, infers what the items are asking, and then responds in like manner to the remaining items without attending to the actual wording of the questions. When this happens, the reversed wording of the question is not noticed by the participant and thus the response to that item will not reflect a participant's true standing on the measured construct. In fact, Schmitt and Stuits found that only 10% of careless respondents are needed to create a distinguishable, negatively-keyed factor. Given this evidence, the removal of the negatively-keyed items from all three scales seemed statistically justified in the present context.

Following the removal of the aforementioned questions from each scale, an exploratory factor analysis utilizing the principal axis factoring model was conducted on the remaining 14 items from the mechanical interests, knowledge, and experiences scales. A varimax rotation was then applied to any extracted factors meeting Kaiser's normalization criterion. The purpose of this analysis was to determine if treating each subscale as representative of a single unique construct (i.e., mechanical interests, mechanical knowledge, and mechanical experiences) was justified or if instead they together described one or more different constructs (i.e., mechanical background, etc.). The results of this procedure are presented in Table 6. Although two factors were

Table 6

Unrotated and Rotated (Varimax) Factor Loading Matrix for Items from the Mechanical Interests, Knowledge and Experiences Subscales

	Unrotated		Rotated	
Items	Factor I	Factor II	Factor I	Factor II
Mechanical Knowledge #1	.64		.60	
Mechanical Knowledge #2	.61		.55	
Mechanical Knowledge #3	.43	.39	.41	
Mechanical Knowledge #5	.66		.52	.41
Mechanical Experiences #1	.74		.60	.43
Mechanical Experiences #2	.78		.70	
Mechanical Experiences #3	.74		.66	.36
Mechanical Experiences #5	.75		.69	.35
Mechanical Experiences #6	.80		.77	.34
Mechanical Interest #2	.65		.32	.62
Mechanical Interest #3	.72		.41	.62
Mechanical Interest #4	.73	.39		.77
Mechanical Interest #5	.71	.33	.33	.71
Mechanical Knowledge #6	.57		.35	.46
Eigenvalue	7.05	1.05	4.04	3.16
% of variance	50.3%	7.5%	28.8%	22.6%

Note. Bold values show the highest factor loading. Loadings below .30 are not presented to ease interpretation.

initially extracted with eigenvalues greater than 1.0, an analysis of the scree plot for the data indicated that only the first factor be retained. However, research (cf. Levonian & Comrey, 1966; Guertin, Guertin, & Ware, 1981; Rummel, 1970) and theoretical rationale (cf. Tinsley & Tinsley, 1987) in the broader literature suggest that in cases where the extraction criteria result in an ambiguous or contradictory statistical conclusion, it is better to overestimate rather than underestimate the number of factors kept for interpretation. Therefore, the decision was made to retain both factors for rotation. Inspection of the loading matrix after rotation revealed that the two factor solution

approximated the criteria for simple structure reasonably well. Items on the mechanical experiences scale seemed to suffer from the highest degree of cross-loading, but in nearly all cases the higher loadings were sufficiently large to justify ascribing the item to a single factor.

Following rotation, the two factors together explained 51.4% of the variance in the measures. Factor I, which accounted for 28.8% of the variance following rotation, was primarily composed of the mechanical knowledge and mechanical experiences items; Factor II accounted for 22.6% of the total variance after rotation and was made up of the remaining mechanical interest items and a single question from the mechanical knowledge scale. In order to determine whether any broad themes emerged that could be used to label and interpret the individual factors, the content of the clustered items was inspected. Items loading on Factor I appeared to represent an individual's behavioral, "outward" engagement with mechanically-related situations and activities, exemplified by items such as "I typically *make repairs* around the house when they are needed by myself rather than ask for help" (Mechanical Experiences #2) and "I am often asked to *show or explain* to others how to operate a piece of mechanical equipment (e.g., run a lawnmower, use a power tool, use a sewing machine, etc.)" (Mechanical Knowledge #1). In keeping with the previous naming convention, this combined factor was labeled *mechanical experience*. The items loading on Factor II, on the other hand, tended to focus on a person's cognitive, "inward" engagement with mechanical incidents. Typified by questions such as "When faced with an object that isn't working properly (such as an appliance or a bicycle), I enjoy trying to *figure out* the causes of the malfunction" (Mechanical Interests #4) and "I try to *develop strategies or techniques* (e.g. trial and

error, working backwards, etc.) for approaching mechanical or technical activities"

(Mechanical Knowledge #6), this factor was most consistent with the previous notion of

mechanical interest.

Table 7 presents the items, descriptive statistics, and reliabilities for the final

mechanical interest and mechanical experience factors. As can be seen in the table, the

newly constructed mechanical experience ($\alpha = .89$) and mechanical interest ($\alpha = .85$)

scales both demonstrated good overall reliability. On the basis of the preceding analyses,

participants' responses on the mechanical interest, knowledge and experience subscales

were combined as shown in Table 7, and the average scale scores used for subsequent

hypothesis testing.

Mechanical Self-Efficacy. There were no negatively keyed items included in the

mechanical self-efficacy measure and thus scale reliabilities could be calculated on the

item responses as recorded. Cronbach's alpha for the 8-item questionnaire was

sufficiently high ($\alpha = .93$), and no improvements to the reliability of the scale could be

achieved by removing any of the items. To determine the underlying structure of the

questionnaire, principal axis factoring was conducted. A single factor solution ($\lambda = 5.437$)

was supported by the analysis which explained 68.0% of the variance in the mechanical

self-efficacy scale. As a result, no changes were made to the measure and participants'

average scale score was used for hypothesis testing.

Gender role identification. No items in either the Masculinity or Femininity

subscale of the BSRI required reverse coding. The 20-item Masculinity subscale

demonstrated a reliability of $\alpha = .86$, whereas the 20-item Femininity subscale achieved

an alpha of $\alpha = .83$. Both of these values are comparable with the reliabilities originally

Table 7

Items, Reliabilities, Means and Standard Deviations (SD) for Mechanical Experience and Mechanical Interest Scales

Items	Reliability[a]	Means	SD
Mechanical Experience	.89		
I am often asked to show or explain to others how to operate a piece of mechanical equipment (e.g., run a lawnmower, use a power tool, use a sewing machine, etc.).	.64	2.76	1.03
I have had much INFORMAL training related to mechanical or technical activities (e.g., being taught how to fix a car, leaky faucet, etc. by a relative or friend).	.59	2.84	1.14
I have had much FORMAL training related to mechanical or technical activities (e.g., a school physics course, vocational/on-the-job training, etc.).	.42	2.15	.89
I can usually figure out what is wrong with an object that is not working correctly.	.61	3.13	.87
When I was growing up, I often helped fix things around the house.	.69	3.00	1.07
I typically make repairs around the house when they are needed by myself rather than ask for help.	.74	2.91	1.05
I have frequently taken part in skilled manual activities (e.g., home improvement, auto repair, sewing projects).	.71	2.73	1.07
I have performed many tasks that require the use of hand tools.	.72	3.26	1.05
I have had many experiences where my mechanical abilities or skills were helpful in fixing a problem (e.g., changing a flat tire, repairing a broken door, etc.).	.78	3.00	1.08
Mechanical Interest	.85		
I enjoy reading magazines or watching television shows about mechanically related topics (e.g., automobiles, new technologies, gadgets, etc.).	.66	3.06	1.09
I enjoy learning about new techniques for performing mechanical or technical activities (e.g., tips on home improvement/repair, auto maintenance, etc.).	.71	3.15	1.01
When faced with an object that isn't working properly (such as an appliance or a bicycle), I enjoy trying to figure out the causes of the malfunction.	.74	3.25	1.06
I like trying to discover how mechanical devices work (through observation, taking things apart, etc.).	.71	3.07	1.01
I try to develop strategies or techniques (e.g. trial and error, working backwards, etc.) for approaching mechanical or technical activities.	.53	3.32	.92

[a]Reliabilities for items are reported as corrected item-total correlations; reliability for scales are reported as Cronbach's alpha

Note. All answers given on a 5-point response scale (1 = *strongly disagree*; 5 = *strongly agree*)

reported by Bem (1974), suggesting that test takers were responding to the items as intended. As indicated earlier, no changes were made to either subscale as a result of the reliability analyses. An average score was thus computed for each subscale and used in further hypothesis testing.

Gender Stereotype Endorsement. Again, no items in the gender stereotype endorsement measure were negatively keyed and thus no reverse coding was necessary. An initial analysis of the 10-item scale revealed acceptable reliability ($\alpha = .84$), and an item level examination of the measure did not suggest removing any questions from the instrument. An exploratory factor analysis of the measure's structure was conducted using the principal axis factoring model. A varimax rotation was applied to any extracted factors meeting Kaiser's normalization criterion. The results of this procedure are presented in Table 8. Two factors with eigenvalues greater than 1.0 were extracted, a decision also supported by examination of the scree plot. The rotated factor structure did not clearly support a simple structure patterning in the data, with items 2, 4, and 6 displaying a fairly high degree of cross-loading.

To determine if a reasonable interpretation of the factor structure could be achieved, an analysis of the item content was conducted. Items 1, 3, 5, 7 and 10, which appeared to load rather cleanly on Factor 1, all began by asking respondents to rate the degree to which they believed "Women are better at _____ than men." Alternatively, items 2, 4, 6, 8 and 9, which either demonstrated significant cross-loadings or strong loadings on Factor 2, asked participants to rate the degree to which they believed "Men are better at _____ than women." Although the results of the factor analysis do not support such an unambiguous interpretation of the data, drawing a distinction between

Table 8

Unrotated and Rotated (Varimax) Factor Loading Matrix for Items from the Gender Stereotype Endorsement Scale

Items	Unrotated		Rotated	
	Factor I	Factor II	Factor I	Factor II
Stereotype Endorsement #1	**.49**		**.54**	
Stereotype Endorsement #3	**.54**	-.38	**.66**	
Stereotype Endorsement #5	**.62**	-.43	**.75**	
Stereotype Endorsement #7	**.67**		**.56**	.37
Stereotype Endorsement #10	**.62**		**.53**	.33
Stereotype Endorsement #2	**.54**		**.41**	.35
Stereotype Endorsement #4	**.62**		**.44**	**.44**
Stereotype Endorsement #6	**.72**		**.53**	.48
Stereotype Endorsement #8	.46	**.56**		**.72**
Stereotype Endorsement #9	**.64**	.53		**.81**
Eigenvalue	4.10	1.46	2.57	1.99
% of variance	41.0%	14.6%	25.7%	19.9%

Note. Bold values show the highest factor loading. Loadings below .30 are not presented to ease interpretation.

these two factors in this manner makes sense conceptually. Originally, the measure was meant to capture the degree to which individuals truly believed there were differences in ability that could be explained by gender in an *absolute* sense; in other words, the gender stereotype endorsement survey was intended to capture an individual's overall appraisal of whether ability differences could be ascribed to differences between men and women. For example, it was believed that a person who responded high to all items on the gender stereotype endorsement scale was someone who wholly endorsed common stereotypic ability differences across gender. However, as suggested by this analysis, it appears that individuals may have been endorsing beliefs about the stereotypic superiority of men and women *separately*. Explained differently, the scale seems to be capturing an individual's

endorsement of stereotypes concerning women *and* their endorsement of stereotypes concerning men.

Based on the preceding discussion, Factor I (which explained 41.0% of the variance before rotation and 25.7% after rotation) can be labeled *female stereotype endorsement*, and Factor II (which explained 14.6% of the variance after rotation and 19.9% after rotation) can be labeled *male stereotype endorsement*. Table 9 presents the items, descriptive statistics, and reliabilities for the final stereotype endorsement scales. As shown, an examination of the 5-item subscale for female stereotype endorsement found an $\alpha = .78$ while the 5-item male stereotype endorsement subscale achieved an $\alpha = .77$, both of which are acceptable levels for the purposes of this research. Additionally, the factor scores from each extracted factor were not significantly correlated ($r = .11$, $p > .05$), thus reaffirming the conclusion that the original gender stereotype endorsement scale was likely capturing two unique facets. Thus, an average scale score was calculated for each individual on female stereotype endorsement and male stereotype endorsement. To examine the effects of these variables in subsequent hypothesis testing, all analyses that incorporated gender stereotype endorsement were run twice—once using female stereotype endorsement and once using male stereotype endorsement.

Mechanical Comprehension. No items in the BMCT require reverse coding. The 68-item test had a reliability of $\alpha = .83$ which is comparable to the median value reported by Bennett (1969) in the original test manual. As indicated earlier, no changes were made to the overall scale as a result of the reliability analyses. Each item on the test was scored correct/incorrect and a final total score was computed for the measure to be used in further hypothesis testing.

Table 9

Items, Reliabilities, Means and Standard Deviations (SD) for Female Stereotype and Male Stereotype Endorsement Scales

Items	Reliability[a]	Means	SD
Female Stereotype Endorsement	.78		
Women have better organization skills than men.	.52	3.14	1.02
Women have greater verbal ability than men.	.52	2.85	1.24
Women have better interpersonal skills than men.	.58	3.13	1.13
Women have better clerical skills (i.e., are good at typing quickly and accurately, identifying mistakes in printed documents, etc.) than men.	.55	2.70	1.19
Women have more creative ability than men.	.56	2.29	1.21
Male Stereotype Endorsement	.77		
Men have better spatial skills than women.	.46	2.90	.96
Men have better math skills than women.	.55	2.40	1.16
Men have better mechanical reasoning skills than women.	.59	3.32	1.10
Men have more overall intellectual ability than women.	.49	1.59	1.00
Men have better analytic reasoning skills than women.	.65	2.07	1.11

[a]Reliabilities for items are reported as corrected item-total correlations; reliability for scales are reported as Cronbach's alpha

Note. All responses are given on a 5-point scale (1 = *not at all true*; 5 = *absolutely true*)

The means, standard deviations and correlations for all study variables are presented in Table 10. Additionally, Table 11 gives a comparison of the means and standard deviations as well as overall effect sizes for each variable across gender. Significant mean differences were found between males and females for all the measured study variables, with a number of the effect sizes approaching one standard deviation. The observed performance discrepancy in mechanical comprehension ($d = 1.08$) was comparable to that reported in the test manual for the BMCT Form S ($d = .8$ to 1.2; Bennett, 1969), which lends strength to the representativeness of the current study's sample in relation to the sample from which the normative statistics of the test were drawn. For the analyses presented below, all continuous predictor variables were centered prior to entry into the regression equations; thus, all beta weights should be interpreted as the effect of the independent variable on the dependent variable when all other variables in the equation are at their respective grand means (Cohen, Cohen, West, & Aiken, 2003)[3].

Hypothesis 1. Hypothesis 1 predicted that individuals with higher levels of mechanical interest and experience would report higher levels of mechanical self-efficacy. To test this hypothesis, mechanical self-efficacy was regressed onto mechanical interest and experience. Table 12 shows that both mechanical interest and mechanical experience predicted significant variance in mechanical self-efficacy ($R^2 = .45$, $F(2, 255) = 105.71$, $p < .001$). Examination of the regression coefficients revealed that as the level of interest ($b = .52$, $t(255) = 4.724$, $p < .001$) and experience ($b = .59$, $t(255) = 6.022$, $p < .001$)

---

[3] The dummy coded gender variable was not centered prior to its inclusion in the regression equations.

Table 10

Means, Standard Deviations and Correlations of Study Variables

Variable	M	SD	1	2	3	4	5	6	7	8	9
1. Mechanical Interest[a]	3.03	.70	(.85)								
2. Mechanical Experience[a]	2.87	.78	**.73****	(.89)							
3. Mechanical Self-efficacy[b]	3.74	1.13	**.61****	**.64****	(.93)						
4. BSRI (M)[b]	5.03	.65	**.34****	**.47****	**.33****	(.86)					
5. BSRI (F)[b]	4.98	.59	**-.13***	**-.14***	**-.16***	.07	(.83)				
6. Female Stereotype Endorsement[a]	2.83	.85	**-.15***	**-.13***	**-.15***	.03	**.28****	(.78)			
7. Male Stereotype Endorsement[a]	2.47	.78	.05	.05	.03	**.13***	-.07	**.55****	(.77)		
8. BMCT	41.39	8.21	**.34****	**.23****	**.40****	.05	**-.26****	**-.32****	-.04	(.83)	
9. Gender[c]	.38	.49	**.42****	**.36****	**.41****	**.13***	**-.43****	**-.28****	**.18****	**.47****	--

Significant correlations are presented in bold; $* = p < .05$, $** = p < .01$; $n = 258$, α reported on diagonal where applicable
[a]Measured on 5-point response scale
[b]Measured on 7-point response scale
[c]Dummy coded variable, Female = 0, Male =1
Note. BSRI (M) = Masculine subscale of Bem Sex Role Inventory; BSRI (F) = Feminine subscale of Bem Sex Role Inventory; BMCT = Bennett Mechanical Comprehension Test

Table 11

Means, Standard Deviations, and Effect Sizes of Study Variables Across Gender

	Females (n = 159)		Males (n = 99)		t^a	d
	M	SD	M	SD		
1. Mechanical Interest	2.80	.63	3.41	.64	7.486***	.96
2. Mechanical Experience	2.65	.71	3.22	.75	6.143***	.78
3. Mechanical Self-efficacy	3.37	1.07	4.33	.97	7.280***	.94
4. BSRI (M)	4.96	.62	5.14	.68	2.171*	.26
5. BSRI (F)	5.18	.52	4.66	.56	7.675***	-.96
6. Female Stereotype Endorsement	3.02	.84	2.53	.77	4.739***	-.61
7. Male Stereotype Endorsement	2.36	.69	2.64	.88	2.878**	.35
8. BMCT	38.35	7.05	46.26	7.60	8.502***	1.08

A positive *d* value indicates that males scored higher than females

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

$^a df = 256$

Note. BSRI (M) = Masculine subscale of Bem Sex Role Inventory; BSRI (F) = Feminine subscale of Bem Sex Role Inventory; BMCT = Bennett Mechanical Comprehension Test

concerning mechanical-related activities increases, one's reported level of mechanical self-efficacy also increases. Thus, this analysis fully supports the predictions made by Hypothesis 1.

Hypothesis 2. To determine whether the relationship between mechanical interest and experience on BMCT performance was mediated by mechanical self-efficacy, the four step regression procedure for testing mediation effects outlined by Baron and Kenny

Table 12

Regression Coefficients for Mechanical Interest and Experience in Predicting Mechanical Self-Efficacy (Hypothesis 1)

	b	β	t^a	R^2
Mechanical Interest	.516	.318	4.724***	
Mechanical Experience	.589	.405	6.022***	.45***

*** $p < .001$

$^a df = 255$

(1986) was used. The results of the analyses are presented in Table 13. In the first step of this procedure, the outcome variable is regressed onto the predictor variable(s) to demonstrate that a direct effect exists which may be mediated. In the current analyses, only the relationship between mechanical interest and BMCT scores was significant ($b = 4.36$, $t(255) = 4.330$, $p < .001$); the regression coefficient for mechanical experience failed to achieve significance ($b = -.44$, $t(255) = .492$, $n.s.$). Step two requires that the predictor must also be related to the proposed mediator. As demonstrated in the analysis for Hypothesis 1, both mechanical interest and mechanical experience were significantly related to mechanical self-efficacy. The third step in the mediation analysis requires that the mediator be related to the outcome when controlling for the effects of the independent variable(s); in the present case, the relationship between mechanical self-efficacy and BMCT scores controlling for mechanical interest and mechanical experience did attain significance ($b = 2.68$, $t(254) = 4.845$, $p < .001$). In the final step, full mediation is said to be achieved if the relationship between the predictor and the outcome drops to zero

Table 13

Test for Mediating Effects of Mechanical Self-Efficacy on Relationship Between Mechanical Interest and BMCT Scores (Hypothesis 2)

DV	Baron & Kenny (1986) Steps	b	R^2
BMCT	1. Mechanical Interest	4.361***	
	Mechanical Experience	-.444	.116***
Mechanical Self-Efficacy	2. Mechanical Interest	.516***	
	Mechanical Experience	.589***	.452***
BMCT	3. Mechanical Self-Efficacy	2.683***	
	4. Mechanical Interest	2.977**	
	Mechanical Experience	-2.024*	.191***

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

when the mediator is introduced into the regression equation; however, in the case that steps one through three are met but the coefficient does not decrease completely to zero, partial mediation is indicated. As seen in Table 13, the regression coefficient for mechanical interest, though attenuated, does not reach zero (b = 2.98, $t(255)$ = 2.956, p < .01), thus supporting a conclusion of partial mediation. In the case of mechanical experience, the regression coefficient increased (b = -2.02, $t(255)$ = 2.190, p < .05).

In summary, according to the steps proposed by Baron and Kenny (1986) for testing mediation, it would appear that only the relationship between mechanical interest and mechanical comprehension is mediated by mechanical self-efficacy—neither the initial (Step 1) nor final (Step 4) criteria for mediation were met in the case of mechanical experience. However, an examination of the pattern of regression coefficients and interrcorrelations among the mediated model's variables actually suggests that a special circumstance termed "inconsistent mediation" may actually be occurring (MacKinnon, Fairchild, & Fritz, 2007). Inconsistent mediation models are models in which at least one mediated effect has a different sign (+/-) than other mediated or direct effects in the model. This is evidenced in the present study if one compares the negative regression coefficient for the path between mechanical experience and BMCT scores in Step 4 (the direct effect; b = -2.024) to the positive product of the mediating path coefficients in Steps 2 and 3 (mechanical experience → mechanical self-efficacy, b = .589, x mechanical self-efficacy → BMCT, b = 2.683, = +1.580).

MacKinnon et al. (2007) point out that, although demonstrating the significance of the relationship between the independent (X) and dependent variable (Y) is important for interpreting the results of the model (i.e., satisfying Step 1 of the Baron & Kenny,

1986, procedure), there can be a number of cases in which the X to Y relation is nonsignificant though mediation still exists. Furthermore, inconsistent mediation models are commonly observed in models such as the present, in which there is more than one mediated effect being tested (mechanical interest → self-efficacy → BMCT; mechanical experience → self-efficacy → BMCT). In this case, the mediator can actually behave as a suppressor variable in the presence of two or more moderately correlated independent variables, thus masking certain expected X to Y relations. As Table 10 shows, the problem of interrcorrelation among independent variables is a definite concern in the presently tested mediation model, in which mechanical experiences and mechanical interests are correlated at $r = .73$. Given the potential for inconsistent mediation models, Kenny (2008) has since revised the original requirement that all four steps of the Baron and Kenny (1986) procedure be met in order to establish mediation; instead, demonstrating support for Steps 2 and 3 is considered evidence enough of mediation. In the present analyses, both mechanical interest and mechanical experience satisfied Steps 2 and 3 in the model, and thus Hypothesis 2 was fully supported.

Hypothesis 3a and 3b. An evaluation of Hypotheses 3a and 3b can be obtained by examining the correlation matrix presented in Table 10. For Hypothesis 3a, femininity was predicted to be negatively correlated with mechanical interest and experience. Evidence was found to support these proposed relationships, with femininity and mechanical interest obtaining a Pearson $r(258) = -.13$, $r^2 = .02$, $p < .05$ while the correlation between femininity and mechanical experience reached an $r(258) = -.14$, $r^2 = .02$, $p < .05$. The analyses also yielded findings in favor of Hypothesis 3b, which stated that masculinity should be positively correlated with both mechanical interest ($r(258)$

69

= .34, r^2 = .12, p < .01) and mechanical experience (r(258) = .47, r^2 = .22, p < .01). Overall then, both Hypothesis 3a and 3b were fully supported by the present study.

Hypothesis 4. Owing to the multidimensional structure observed in the gender stereotype endorsement measure employed by this study, an analysis of Hypothesis 4 as originally described was not pursued. Instead, each proposed relationship was analyzed twice, once utilizing an individual's level of female stereotype endorsement and again with one's level of male stereotype endorsement.

The first portion of Hypothesis 4 proposed a significant interaction between stereotype endorsement and femininity on mechanical self-efficacy; the top half of Table 14 displays the results of this analysis. The regression model including femininity, female stereotype endorsement and the interaction term accounted for a significant portion of the variance in mechanical self-efficacy (R^2 = .05, F(3, 254) = 3.970, p < .01). However, an examination of the regression coefficients revealed that only the main effect of femininity was a significant predictor of mechanical self-efficacy (b = -.26, t(254) = 2.126, p < .05), indicating that higher levels of femininity predicted lower levels of mechanical self-efficacy. The model including femininity, male stereotype endorsement and the interaction term failed to account for a significant portion of the variance in mechanical self-efficacy (R^2 = .03, F(3, 254) = 2.483, *n.s.*).

It was also predicted that reported mechanical self-efficacy would not be affected by one's level of stereotype endorsement for masculine identified individuals; the bottom half of Table 14 presents the findings from this analysis. The regression model including masculinity, female stereotype endorsement and the interaction term accounted for a significant portion of the variance in mechanical self-efficacy (R^2 = .14, F(3, 254) =

Table 14

Effects of Gender Role Identification and Gender Stereotype Endorsement on Mechanical Self-Efficacy (Hypothesis 4)

Variable	b	β	R^2
Femininity (F)	-.261*	-.136	
Female Stereotype Endorsement (FE)	-.136	-.102	
F x FE	-.201	-.090	.045**
Femininity (F)	-.293*	-.153	
Male Stereotype Endorsement (ME)	.012	.009	
F x ME	-.112	-.051	.028
Masculinity (M)	.595***	.341	
Female Stereotype Endorsement (FE)	-.202*	-.151	
M x FE	-.141	-.067	.138***
Masculinity (M)	.562***	.323	
Male Stereotype Endorsement (ME)	-.035	-.024	
M x ME	.109	.053	.112***

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

13.526, $p < .001$). An examination of the regression coefficients indicated that both the main effects of masculinity ($b = .60$, $t(254) = 5.831$, $p < .001$) and female stereotype endorsement ($b = -.20$, $t(254) = 2.581$, $p < .05$) were significant predictors of mechanical self-efficacy, though the interaction effect failed to reach significance. The main effects show that higher levels of masculinity predicted higher levels of mechanical self-efficacy, whereas greater endorsement of female stereotypes predicted lower levels of mechanical self-efficacy. The second regression model, which included masculinity, male stereotype endorsement and the interaction term, also accounted for a significant portion of the variance in reported levels of mechanical self-efficacy ($R^2 = .11$, $F(3, 254) = 10.642$, $p < .001$). Only the coefficient for masculinity ($b = .562$, $t(254) = 5.332$, $p < .001$) was significant, indicating that more masculine individuals reported higher levels of mechanical self-efficacy overall.

Taken together, these findings only partially support the predictions made by Hypothesis 4. As predicted, no significant interaction between masculinity and stereotype endorsement (either male or female) emerged; however, the significant interaction proposed to exist between femininity and stereotype endorsement (male or female) was not observed.

Hypothesis 5. Hypothesis 5 proposed a significant three-way interaction between gender, gender role identification, and gender stereotype endorsement on BMCT performance, with the prediction suggesting a differential effect of stereotype endorsement. As was the case in Hypothesis 4, Hypothesis 5 was not tested as originally described because of the two-dimensional structure of the stereotype endorsement scale. Instead, four regression models were estimated, one for each combination of the 2 gender role identification (feminine, masculine) x 2 gender stereotype endorsement (female stereotype endorsement, male stereotype endorsement) possibilities.

The first regression model estimated included gender, femininity, female stereotype endorsement, all two-way interactions, and the single three-way interaction in the equation. Model 1 in Table 15 presents the results of this analysis. The complete model predicted a significant amount of variance in BMCT performance ($R^2 = .27$, $F(7, 250) = 13.222$, $p < .001$). An examination of the regression coefficients revealed that only the main effect for gender was a significant predictor of BMCT performance ($b = 6.60$, $t(250) = 6.227$, $p < .001$), indicating that males tended to score higher on the test than females. It seems noteworthy to point out that both female stereotype endorsement and the three-way interaction term in this model did attain significance at the $p < .10$ level; however, as neither achieved statistical significance according to conventional

72

Table 15

Effects of Gender, Gender Role Identification and Gender Stereotype Endorsement on BMCT scores (Hypothesis 5)

Model	Variable	b	β	R^2
Model 1	Gender (G)[a]	6.603***	.392	
	Femininity (F)	-.543	-.039	
	Female Stereotype Endorsement (FE)	-1.397[†]	-.144	
	G x F	.931	.043	
	G x FE	-.387	-.023	
	F x FE	-1.028	-.063	
	G x F x FE	3.666[†]	.136	.270***
Model 2	Gender (G)[a]	7.697***	.457	
	Femininity (F)	-1.121	-.080	
	Male Stereotype Endorsement (ME)	-.764	-.073	
	G x F	-.427	-.020	
	G x ME	-.621	-.042	
	F x ME	-.408	-.025	
	G x F x ME	1.828	.084	.245***
Model 3	Gender (G)[a]	6.806***	.404	
	Masculinity (M)	.134	.011	
	Female Stereotype Endorsement (FE)	-1.714*	-.177	
	G x M	-.275	-.014	
	G x FE	-.703	-.043	
	M x FE	-.336	-.022	
	G x M x FE	-1.108	-.044	.264***
Model 4	Gender (G)[a]	8.210***	.487	
	Masculinity (M)	-.141	-.011	
	Male Stereotype Endorsement (ME)	-.947	-.090	
	G x M	.478	.025	
	G x ME	-1.046	-.070	
	M x ME	1.121	.075	
	G x M x ME	-.389	-.020	.242***

* $p < .05$, *** $p < .001$, [†] $p < .010$
[a]Dummy coded variable, Female = 0, Male = 1

73

standards, further examination of this finding is not presented here (but see the Additional

Analyses section).

The second model regressed BMCT scores on gender, femininity, male stereotype

endorsement, all possible two-way interactions and the single three-way interaction. As

shown in Table 15 (Model 2), this equation also accounted for a significant amount of

variance in BMCT scores ($R^2 = .25$, $F(7, 250) = 11.608$, $p < .001$). Again, however, only

the regression coefficient for gender ($b = 7.70$, $t(250) = 7.280$, $p < .001$) attained

significance, indicating that males outperformed females on the BMCT.

Model 3 in Table 15 presents the results of the regression equation predicting

BMCT performance using gender, masculinity, female stereotype endorsement, all

possible two-way interactions and the single three-way interaction. The model predicted

a significant portion of the variance in BMCT scores ($R^2 = .26$, $F(7, 250) = 12.826$, p

$< .001$), with the main effects for both gender ($b = 6.81$, $t(250) = 6.947$, $p < .001$) and

female stereotype endorsement ($b = -1.71$, $t(250) = 2.510$, $p < .05$) reaching significance.

The regression coefficient for gender indicated that males tended to score higher than

females on the BMCT; the coefficient for female stereotype endorsement indicated that

individuals who more strongly endorsed stereotypes about females tended to perform

more poorly on the BMCT.

The final regression model (Model 4, Table 15) included gender, masculinity,

male stereotype endorsement, all possible two-way interactions and the single three-way

interaction. The model again predicted significant variance in performance on the BMCT

($R^2 = .24$, $F(7, 250) = 11.376$, $p < .001$). Examination of the regression weights revealed

that gender was the only significant gender predictor in the equation ($b = 8.21$, $t(250) =$

8.557, $p < .001$), showing that males tended to achieve higher scores on the BMCT compared to females.

In summary, gender was the only consistently significant predictor of BMCT performance across all the tested regression models. However, female stereotype endorsement did attain significance in Model 3 and was marginally significant in Model 1, suggesting that individuals who more strongly endorsed stereotypes about females generally had lower mechanical comprehension scores. Finally, although the three-way interaction in Model 1 approached significance, the hypothesized interaction between gender, gender role identification and gender stereotype endorsement failed to achieve significance in any of the tested models. Taken together, these results do not support Hypothesis 5.

Hypothesis 6a, 6b, 6c, and 6d. Hypotheses 6a through 6d sought to explicitly test the conclusions published by Antill and Cunningham (1982), which stated that masculine males were the top scorers on tests of mechanical comprehension, followed by feminine males, masculine females and, lastly, feminine females. Although the median-split procedure used by Antill and Cunningham to produce their findings is a common one in the social sciences, many methodologists discourage the use of the technique because of the unnecessary restrictions and limitations it places on subsequent statistical analyses. Specifically, forcing dichotomization (or any similarly "arbitrary" grouping strategy) is problematic because it capitalizes on sample specific data characteristics, requires perfectly reliable measuring instruments in order to correctly classify individuals at or around the selected cutoff scores, can generate spurious results, and substantially

decrease statistical power (Cohen, 1983; Hunter & Schmidt, 1990; Aiken & West, 1991; Maxwell & Delaney, 1993).

To circumvent these issues, moderated hierarchical regression is often recommended as an alternative and preferred statistical procedure (Aiken & West, 1991). In this technique, the dependent variable is first regressed onto the main effects of the two independent variables. In the next step, the moderator variable (which is simply the cross-product of the two independent variables) is introduced into the equation; if the moderator variable predicts significant variance above and beyond the two main effects, an interactive effect is said to be present in the data and a simple slopes analysis can be performed to determine the pattern of the relationship. Therefore, the same predictions tested with a median split procedure can be performed without suffering from the aforementioned deficiencies. As such, the present analysis was conducted using moderated hierarchical regression.

In the present case, because the four gender role categories identified by Spence et al. (1975) and analyzed by Antill and Cunningham (1982) were constructed by combining information concerning one's level of masculinity *and* femininity, there were actually three independent variables to enter into the hierarchical moderated regression equation—gender, masculinity and femininity. To test Hypotheses 6a-6d, the main effects for gender, masculinity, and femininity were entered into the regression model first, followed by all possible two-way interactions, and lastly the three-way interaction term. Support for the predicted gender by gender role interaction is indicated if the three-way interaction term attains significance and further decomposition of the interaction into simple slopes support the directional predictions made by the hypotheses. Table 18

presents the regression coefficients and beta weights of this full regression model. The

results of the analysis show that the three-way interaction term failed to achieve

significance ($b = -2.98$, $t(250) = 1.213$, $p = .23$); gender was once again the only

significant predictor of BMCT performance ($b = 7.78$, $t(250) = 7.197$, $p < .001$).

In summary, the present study was unable to produce significant findings to

support the results reported by Antill and Cunningham (1982) using hierarchical

moderated regression. Furthermore, even when analyzing the present data with the same

median split procedure originally used by Antil and Cunningham, evidence was not

found to support the original findings (see Appendix H for a detailed presentation of

these analyses). In both cases, gender was the only variable found to predict one's scores

on the BMCT, with both gender role identification and the accompanying interactions

failing to predict additional variance in mechanical comprehension test performance.

Hence, no evidence was found to support the predictions made by Hypotheses 6a-6d.

Table 16
Effects of Gender and Gender Role Identification on
BMCT Performance (Hypotheses 6a – 6d)

Variable	b	β	t[a]	R^2
Gender (G)[b]	7.776	.461	7.197***	
Femininity (F)	-1.151	-.083	1.019	
Masculinity (M)	.086	.007	.088	
G x F	.620	.029	.355	
G x M	-1.552	-.080	.954	
M x F	-1.034	-.052	.607	
G x F x M	-2.976	-.110	1.213	.491***

*** $p < .001$
[a]$df = 250$
[b]Dummy coded variable, Female = 0, Male = 1

Between-Gender Tests. One of the primary goals of the present study was to specifically examine and identify potential explanations for the sizable discrepancies in mechanical ability test performance commonly observed between men and women. However, large significant gender differences were observed not only for scores on the BMCT, but nearly every variable measured in the present study (see Table 11). This suggests that the variables captured here might offer additional insight into the mechanisms that generate the noted performance discrepancies between men and women if their effects are examined while controlling for participant sex. As such, Hypotheses 1 through 4 (which did not previously include participant sex in the analyses) were run again, this time partialling out the effects of gender to determine if the observed relationships still functioned similarly.

To reanalyze Hypothesis 1, mechanical self-efficacy was regressed onto mechanical interest and experience, though this time gender was first entered as a control variable into the regression model (Table 17). In the first step, gender explained a significant amount of the variance in mechanical self-efficacy ($R^2 = .172$, $F(1, 256) = 53.00$, $p < .001$); the addition of mechanical interest and mechanical experience into the equation explained an additional 30% of the variance in mechanical self-efficacy ($\Delta R^2 = .303$, $\Delta F(2, 254) = 73.08$, $p < .001$). The regression coefficients again showed that, even when controlling for gender, both mechanical interest ($b = .42$, $t(254) = 3.821$, $p < .001$) and mechanical experience ($b = .56$, $t(254) = 5.846$, $p < .001$) were significant positive predictors of one's reported level of mechanical self-efficacy above and beyond gender ($b = .38$, $t(254) = 3.267$, $p < .001$). Thus the propositions presented in Hypothesis

1 were supported: regardless of a person's sex, greater levels of mechanical interest and experience predict greater levels of mechanical self-efficacy.

Table 18 presents the results for Hypothesis 2 controlling for gender. Similar to the findings reported earlier, mechanical self-efficacy was still found to partially mediate the relationship between mechanical interest and mechanical experience with performance on the BMCT. Furthermore, the regression coefficient for mechanical interests in the final step of the mediation analysis dropped even more substantially than in the previous analysis, so much so that it was no longer a significant predictor of mechanical self-efficacy above and beyond gender and mechanical experiences. Together, this finding lends further support to the conclusion posited in Hypothesis 2 that mechanical self-efficacy is a significant mediator of the relationship between one's mechanical interest and experience and their mechanical aptitude.

To reevaluate Hypotheses 3a and 3b, partial correlation coefficients were calculated for mechanical interest and experience with scores on the BSRI masculinity and femininity subscales excluding the variance accounted for by gender. Contrary to the previous finding, when gender variance was removed from the correlation coefficient,

Table 17
Regression Coefficients for Mechanical Interest and Experience in Predicting Mechanical Self-Efficacy Beyond Gender (Hypothesis 1)

Variable	b	β	R^2	ΔR^2
(Gender)	.961	.414***	.172	.172***
(Gender)	.382	.165***		
Mechanical Interest	.424	.261***		
Mechanical Experience	.563	.387***	.474	.303***

** $p < .01$, *** $p < .001$
[a]$df = 254$
Note. Gender was used as a control variable in the regression equation

Table 18

Mediating Effects of Mechanical Self-Efficacy on Relationship Between Mechanical Interest and Experience and BMCT Scores Controlling for Gender (Hypothesis 2)

DV	Step	b	R^2	ΔR^2
BMCT	(Gender)	7.910***	.220	
	1. (Gender)	6.775***		
	Mechanical Interest	2.724**		
	Mechanical Experience	-.902	.248	.028**
Mechanical Self-efficacy	(Gender)	.961***	.172	
	2. (Gender)	.382***		
	Mechanical Interest	.424***		
	Mechanical Experience	.563***	.474	.303***
BMCT	(Gender)	7.910***	.220	
	3. (Gender)	6.160***		
	Mechanical Self-efficacy	1.821***	.272	.052***
	4. (Gender)	5.991***		
	Mechanical Self-efficacy	2.051***		
	Mechanical Interest	1.856		
	Mechanical Experience	-2.057*	.290	.017*

*$p < .05$, ** $p < .01$, *** $p < .001$
Note. BMCT = Bennett Mechanical Comprehension Test
Gender was used as a control variable in the regression equations

Hypothesis 3a was no longer supported; rather than demonstrating a significant negative relationship, near zero partial correlations were observed between femininity and mechanical interest ($r_{\text{Fem*Interest•Gender}} = .06$, *n.s.*) as well as femininity and mechanical experience ($r_{\text{Fem*Experience•Gender}} = .02$, *n.s.*). However, significant positive coefficients were still observed between masculinity and mechanical interest ($r_{\text{Masc*Interest•Gender}} = .32$, *p*

< .001) and experience ($r_{\text{Masc*Experience*Gender}}$ = .46, p < .001), thus fully supporting Hypothesis 3b.

The interaction between gender role identification and gender stereotype endorsement (male and female) on reported levels of mechanical self-efficacy above and beyond the effects of gender was examined for the reevaluation of Hypothesis 4 (Table 19). Although the main effects for masculinity were still strong significant predictors of mechanical self-efficacy even after controlling for gender, no other predictors in the equation aside from participant sex achieved significance. Again, none of the gender role by gender stereotype endorsement interactions attained significance. Thus, the predicted interaction proposed in Hypothesis 4 remained unsupported by the present analyses.

Male-Female Stereotype Endorsement. The separation of the gender stereotype endorsement measure into subscales that captured beliefs of female and male stereotypes uniquely opened up the possibility for examining possible interactions between these distinct constructs. As there was no a priori basis on which to make predictions concerning their interactive effects, the following examinations were undertaken as an exploratory investigation of the influence differential endorsements of male versus female stereotypes by males and females have on reported levels of mechanical self-efficacy and mechanical comprehension test performance. Thus, two regression models were conducted, both of which included gender, female stereotype endorsement and male stereotype endorsement as predictors in their respective equations.

The full model in which mechanical self-efficacy was regressed onto gender, female stereotype endorsement and male stereotype endorsement predicted a significant amount of variance in the dependent variable (R^2 = .205, $F(2, 249)$ = 21.483, p < .001).

81

Table 19

Effects of Gender Role Identification and Gender Stereotype Endorsement on Mechanical Self-Efficacy (Hypothesis 4)

Variable	b	β	R^2	ΔR^2
(Gender)	.961***	.414	.172	
(Gender)	.971***	.418		
Femininity (F)	.046	.024		
Female Stereotype Endorsement (FE)	-.036	-.027		
F x FE	-.210	-.094	.182	.010
(Gender)	.961***	.414	.172	
(Gender)	.999***	.431		
Femininity (F)	.047	.025		
Male Stereotype Endorsement (ME)	-.070	-.048		
F x ME	-.038	-.017	.174	.002
(Gender)	.961***	.414	.172	
(Gender)	.832***	.358		
Masculinity (M)	.502***	.288		
Female Stereotype Endorsement (FE)	-.066	-.049		
M x FE	-.100	-.047	.253	.081***
(Gender)	.961***	.414	.172	
(Gender)	.901***	.388		
Masculinity (M)	.499***	.286		
Male Stereotype Endorsement (ME)	-.114	-.079		
M x ME	.020	.010	.254	.082***

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

However, the main effect of gender was the only significant predictor of mechanical self-efficacy ($b = 1.09$, $t(249) = 5.902$, $p < .001$), indicating that males tended to report greater levels of mechanical self-efficacy. Additionally, the main effect of male stereotype endorsement came very close to reaching statistical significance ($b = -.34$, $t(249) = 1.947$, $p = .053$), indicating that individuals who did not believe that stereotypes about males were true tended to have higher levels of reported mechanical self-efficacy. Nevertheless,

in the present model, the differential endorsement of female or male stereotypes did not appear to explain differences in one's reported confidence in their ability to perform mechanical tasks.

In the second regression equation, a significant amount of the observed variance in BMCT performance was explained by the full model including gender, female stereotype endorsement, male stereotype endorsement, and their interactions ($R^2 = .283$, $F(1, 250) = 35.088$, $p < .001$). Here, the main effect of gender was again significant ($b = 8.17$, $t(250) = 6.527$, $p < .001$) indicating that men tended to score higher than women on the test. In addition, the main effect of female stereotype endorsement was significant as well ($b = -2.17$, $t(250) = 2.129$, $p < .05$), demonstrating that individuals who more strongly endorsed stereotypes about women tended to perform worse on the test of mechanical comprehension. However, both of these main effects were qualified by a significant three-way interaction between gender, female stereotype endorsement and male stereotype endorsement ($b = -2.92$, $t(250) = 2.006$, $p < .05$).

To examine the pattern of the significant three-way interaction, the two-way interaction between female and male stereotype endorsement was examined for males and females separately (Figure 7). For female participants, only female stereotype endorsement was a significant predictor of BMCT performance ($b = -2.17$, $t(155) = 2.160$, $p < .05$), indicating that women who more strongly endorsed female stereotypes tended to score lower on the test of mechanical comprehension regardless of their beliefs in male stereotypes. In the case of male participants, only the two-way interaction term achieved significance ($b = -2.61$, $t(95) = 2.487$, $p < .05$). As can be seen in Figure 7, the pattern of the data shows that for men who did not strongly endorse male stereotypes, female
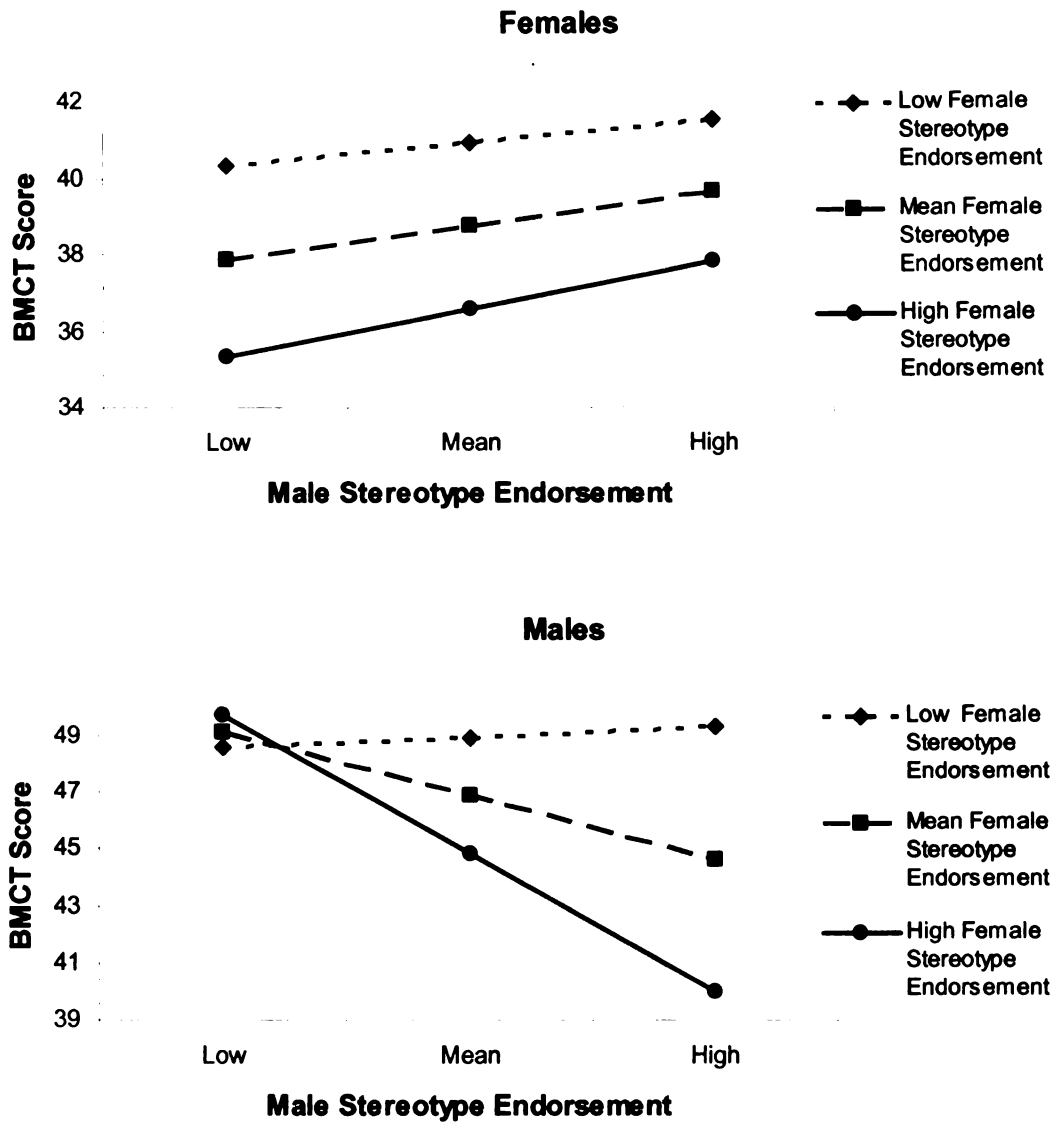
Females

Males

Figure 7. Two-way interaction between female stereotype endorsement and male stereotype endorsement on BMCT performance for females and males

stereotype endorsement made little difference in their overall BMCT performance. However, male participants' belief in female stereotypes greatly affected test performance when they also strongly endorsed male stereotypes; in this case, men who felt more strongly that female stereotypes are true performed significantly worse than those who did not hold such strong stereotypical beliefs about women.

In sum, it appears that female and male stereotype endorsement had little impact on an individual's level of mechanical self-efficacy, with males still reporting the greatest levels of confidence in their mechanical proficiency. On the other hand, these beliefs seemed to have a greater impact on one's performance on the mechanical ability test. Males and those who did not strongly endorse stereotypes about women tended to perform better overall on the test of mechanical comprehension; however, the performance of males was also dependent upon the degree to which they endorsed stereotypes about both men and women.

Hypothesis 5 Follow-up. As mentioned earlier, the model predicted in Hypothesis 5 in which performance on the BMCT was regressed onto gender, femininity and female stereotype endorsement found that both the main effect for female stereotype endorsement ($b = -1.38$, $t(250) = 1.888$, $p = .06$) and the three-way interaction term ($b = 3.67$, $t(250) = 1.671$, $p = .10$) approached, though did not obtain, conventional levels of statistical significance (Table 15). While considered poor practice to interpret these coefficients as "meaningful" findings based on commonly accepted statistical standards, given the somewhat exploratory nature of the hypothesis analysis a cursory examination of their findings seems warranted. Thus the results of those analyses will be presented below, with full knowledge that they should not be interpreted as significant.

The simple main effect of female stereotype endorsement suggests that individuals who more strongly endorse stereotypes about females tended to perform worse on the BMCT. To examine the three-way interaction, the two-way interaction between femininity and female stereotype endorsement was examined across females and males (Figure 8). For females, the two-way interaction term ($b = -1.03$, $t(155) = .893$,

n.s.) indicated that for women who only weakly endorsed female stereotypes, reported identification with the feminine gender role made little difference on BMCT performance; however, for women who strongly endorsed female stereotypes, low feminine-identified women performed slightly better on the BMCT than high-identified women. For males, this pattern was quite different. The two way interaction term ($b = 2.64$, $t(155) = 1.374$, *n.s.*) indicated that when endorsement of female stereotypes was weak, low feminine-identified males outperformed high-feminine identified males on the mechanical comprehension test. However, and perhaps contrary to what might be expected, for those who strongly endorsed female stereotypes, high feminine-identified males achieved higher BMCT scores than low feminine-identified males.
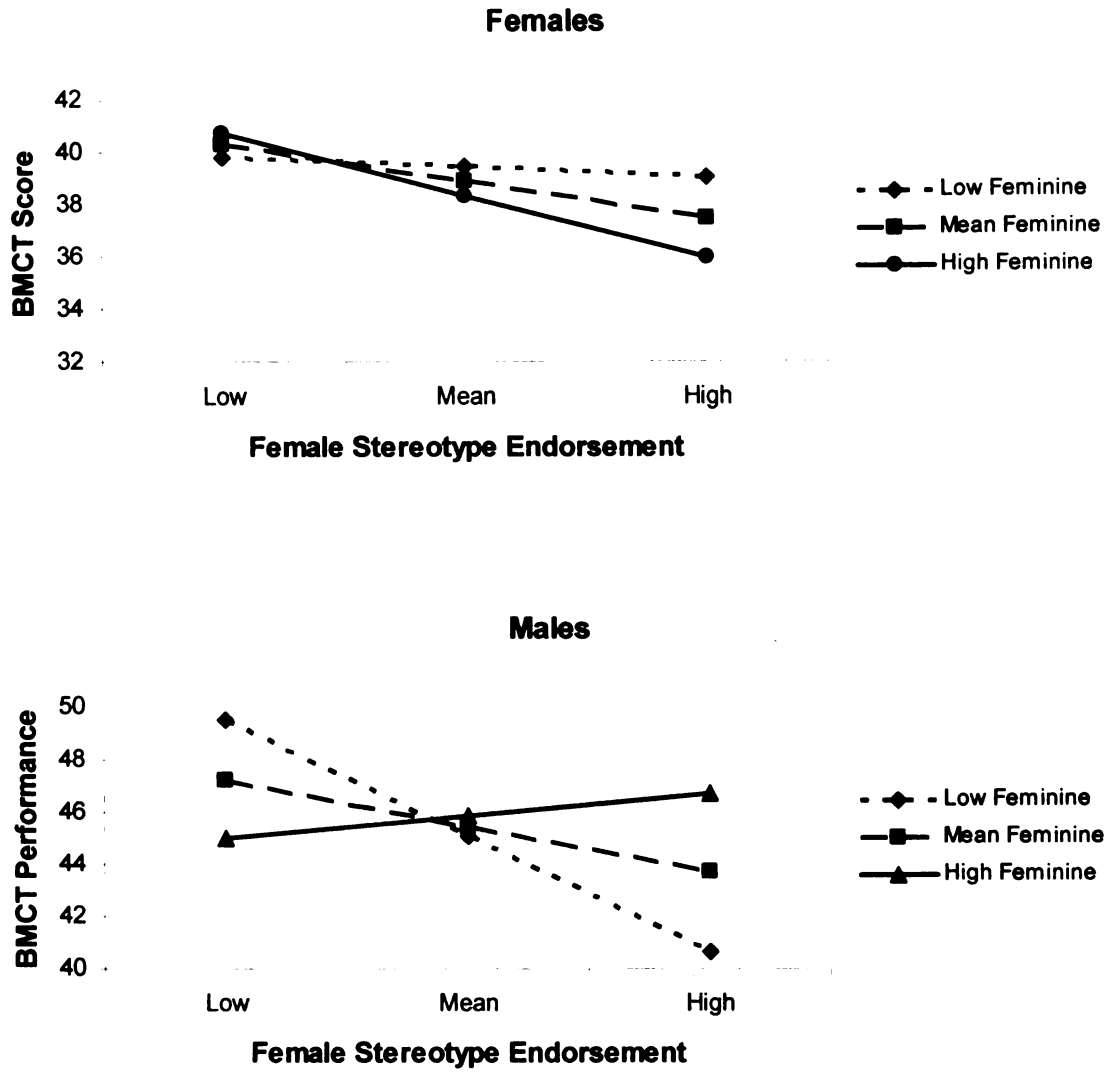
Figure 8. Two-way interaction between feminine gender role identification and female stereotype endorsement on BMCT performance for females and males

DISCUSSION

The purpose of the present study was to examine the influence of key individual difference variables hypothesized to predict the noted discrepancies in mechanical comprehension across gender. Table 20 presents a formal summary of the hypotheses pursued in this experiment and the subsequent conclusions supported by the data. The pattern of results revealed that both mechanical interest and mechanical experiences were fairly strong predictors of mechanical self-efficacy for both males and females, which was itself a significant predictor of performance on the BMCT. Furthermore, an individual's identification with male and female gender role characteristics, and masculinity in particular, was generally indicative of these important mechanical background variables. Finally, the investigation of gender stereotype endorsement pursued by this study showed that personal beliefs about the truthfulness of male and female stereotypes did appear to influence performance on the mechanical comprehension test to some degree, a finding not previously hypothesized or found within the research literature.

In an attempt to better interpret and frame the present findings, the following discussion will broadly focus on an evaluation of the validity of the proposed gender differences model. More specifically, this section's primary focus is on examining which variables from the model appeared to operate in the manner predicted and which did not, and subsequently how these findings contribute to an understanding of specific cognitive abilities and the oft observed male-female differences that accompany their measurement. Attention is also directed towards promoting the applicability of the theoretical/empirical approach adopted in the current study whereby an individual differences framework is

Table 20
Hypothesis Summary

Hypothesis	Result
Hypothesis 1: Individuals with more interest in, greater knowledge of, and more experiences with mechanically-related subject matter will report higher ratings of mechanical self-efficacy.	*Supported*
Hypothesis 2: The relationship between mechanical interests, knowledge and experiences and performance on a test of mechanical comprehension will be mediated by mechanical self-efficacy.	*Supported*
Hypothesis 3a: Individuals who identify more with a feminine gender role will have less interest in, less general knowledge of, and fewer experiences with mechanically-related subject matter.	*Supported**
Hypothesis 3b: Individuals who identify more with a masculine gender role will have more interest in, greater general knowledge of, and more experiences with mechanically-related subject matter.	*Supported*
Hypothesis 4: A significant interaction between gender role identification and gender stereotype endorsement will emerge such that feminine individuals who more strongly endorse gender stereotypes will report lower levels of mechanical self-efficacy than those who do not endorse gender stereotypes. The interaction will not be observed for masculine identified individuals.	Not supported
Hypothesis 5: There will be a three-way interaction between gender, gender role identification, and gender stereotype endorsement such that women who are feminine (or masculine) identified and endorse gender stereotypes will score lower than feminine (or masculine) identified women who do not endorse gender stereotypes; this pattern will not be evidenced for males.	Not supported
Hypothesis 6a: Males who identify with a masculine gender role will score the highest on a test of mechanical comprehension.	Not supported
Hypothesis 6b: Males who identify with a feminine gender role will score lower on a test of mechanical comprehension than masculine males, but higher than masculine females.	Not supported
Hypothesis 6c: Females who identify with a masculine gender role will score higher on a test of mechanical comprehension than feminine females, but lower than feminine males.	Not supported
Hypothesis 6d: Females who identify with a feminine gender role will score the lowest on a test of mechanical comprehension.	Not supported

*Not supported above and beyond the effect of gender

used to model male-female discrepancies in mechanical ability rather than simply ascribing this disparity to sex. Lastly, the discussion will conclude with a presentation of the study's limitations and advice/lessons learned for future related research inquiries.

As stated earlier, the functional model of male-female differences in mechanical comprehension proposed by the current study (Figure 2) was designed to examine the influence of individual-level, psychological/cognitive variables as predictors of performance on mechanical ability tests. Additionally, these individual difference variables were specifically chosen to capture two distinct characteristics of the participant thought to be important to the particular cognitive domain at hand. The first such category broadly focused on the *gender-related* characteristics of the individual (i.e., gender role identification and gender stereotype endorsement) while the second class of variables could be classified as the *contextually-related* characteristics of the individual relative to the testing scenario (i.e., mechanical interests, experiences and self-efficacy). Although the case has already been made for why these particular variables were selected as important indicators of the gender-related or contextually-related characteristics of an individual, there is little question that any number of additional variables could have been used to represent these psychological aspects of the male and female test taker. Limitations of content validity aside, however, reviewing the results of this study by examining the relative impacts of gender-related and contextually-related characteristics on mechanical ability performance offers a useful and meaningful frame of reference for interpreting the present findings and their contribution to the research literature at large.

Gender-related characteristics. The first variable considered among the class of gender-related characteristics was gender role identification. As predicted in Hypotheses 3a and 3b, gender role was significantly related to mechanical interests and experiences. However, contrary to what previous research and theory implied (e.g., Bem, 1974; Nash,

1975; 1979; Spence et al., 1975; Antill & Cunningham, 1982), no evidence was found to support any direct or interactive effects involving masculinity or femininity on mechanical ability (Hypothesis 5, 6a-6d). This non-significant finding was one of the more surprising results of the study given the nearly identical operational definition and measurement approach adopted by the present experiment relative to similar past studies that did find evidence of such a relationship. While it is still somewhat unclear why this inconsistency was observed, there is an evolving body of research led by contemporary gender role researchers that proffers one explanation (e.g., Signorella & Jamison, 1986; Hamilton, 1995; Ritter, 2004). Although their contentions are explicated in greater detail below, the overall theme of these works posits that the validity of the presumed gender role-cognitive performance relationship may no longer be appropriate because of significant fundamental changes in individuals' understanding of what it means to be "masculine" and "feminine." This sociocultural shift, the authors contend, has made the process of identifying one's gender role orientation a much more ambiguous task, which affects the development of a person's overall self-concept and is ultimately believed to influence the attainment of cognitive proficiencies and aptitudes.

Originally, the theory behind the gender role-cognitive ability relationship is most notably attributed to Nash (1979), who posited that individuals, regardless of sex, perform best on a cognitive task when their level of masculinity and femininity (i.e., their gender role identification) is consistent with the gender stereotyping present in the cognitive task at hand. Stated more simply, Nash's (1979) gender role hypothesis claims that masculine individuals should be more capable of successfully completing traditionally masculine (as opposed to feminine) cognitive activities and feminine

91

individuals more capable of successfully completing traditionally feminine (as opposed to masculine) cognitive activities[4]. Implied in this theory is that higher cognitive performance is achieved because the individual possesses or otherwise "gains" some real or perceived advantage because of the compatibility between their gender role characteristics and the task at hand. For example, Massa, Mayer and Bohon (2005) suggest that the benefit is motivational. Here, the authors propose that individuals exert more effort and thus perform better on a cognitive task if they believe it taps a knowledge, skill or ability ("This is a masculine/feminine task") consistent with their personal gender role orientation ("I am masculine/feminine") because it reinforces their perception of the masculinity/femininity inherent in their self-concept[5]. Regardless of the mechanism through which performance is enhanced, however, Nash's (1979) theory ultimately suggests that gender role identification may be an important component of some individuals' cognitive ability performance.

Although numerous examples may be found in the broader research literature that support Nash's (1979) general supposition (e.g., Bernard, Boyle, & Jackling, 1990; Newcombe & Dubas, 1992), an increasingly large number of contradictory findings are beginning to accumulate as well (e.g., Hamilton, 1995; Ritter, 2004) in which the link between gender role and various cognitive ability tasks is either non-existent (as was found in the present study) or exactly the opposite of what would be expected (i.e.,

---

[4] As noted previously (cf. Eccles, 1987), this reasoning is based primarily on social cognitive theories of development and socialization, in which individuals pursue various experiences and activities specific to a given gender role on the basis of the intrinsic and extrinsic reinforcement they receive from them. These pursuits, in turn, should have a subsequently greater influence on the specific proficiencies one develops. This is precisely the framework adopted by the present study's model.

[5] Although the authors did not directly test the motivational aspect of their hypothesis, they did find that whereas masculine women performed significantly better on a spatial ability test when they were told it measured spatial ability, the more feminine identified women performed better on the exact same test when instructed the measure actually captured one's level of empathy. Unfortunately, this pattern of results was not replicated for men, perhaps suggesting a more complex phenomenon than originally proposed.

feminine individuals outperforming masculine individuals on a masculine-oriented task).

A meta-analysis conducted by Signorella and Jamison (1986) offers a particularly poignant example of this empirical ambiguity. Across the meta-analyzed studies, the authors noted that higher masculinity and lower femininity tended to be associated with better performance on mathematical and spatial ability tasks (abilities stereotypically associated with masculinity), a finding which would seem to support Nash's (1979) theory. Unexpectedly, however, this pattern did not hold consistently across gender; in fact, their results suggest that adolescent boys who identified more closely with the feminine gender role actually tended to exhibit *higher* mathematical and spatial ability scores. Furthermore, gender role identification did not appear indicative at all of performance on verbal tasks (a stereotypically feminine-associated ability). Taken together, although these results do not wholly "disprove" Nash's (1979) gender role hypothesis, they do raise some interesting questions concerning the premises on which the theory is based and the subsequent interpretation of the present research's findings.

At its core, Nash's (1979) gender role hypothesis rests on the assumption that individuals are capable of successfully performing two key processes: 1) identifying the "amount" of masculinity and femininity one attributes to their self-concept and thus their overall gender role orientation, and 2) classifying a particular cognitive domain as stereotypically masculine or feminine. A breakdown in either of these premises, then, might account for the non-significant relationship observed between gender role and cognitive performance observed in the present study's model. It seems unlikely in the current research, however, that Nash's (1979) second criterion was violated. As was pointed out earlier, the association of a given cognitive ability and related tasks with a

specific gender role is largely dictated by cultural norms (Eccles, 1987; Williams & Best, 1990; Cleveland et al., 2000; Wood & Eagly, 2002). Through socialization processes and experiential learning within a given culture, individuals come to understand and learn what aptitudes and related activities are typically perceived as masculine and which are typically perceived as feminine. For example, Huston (1983) and Signorella and Vegaga (1984) contend that by adolescence, the majority of Americans have come to understand that spatial, mechanical and mathematical skills are associated with masculinity, while verbal skills come to be considered feminine qualities (similar claims were empirically investigated and supported in earlier works by Spence et al., 1975). As there is nothing to suggest that the sample used in the current research spanned multiple cultural groups within which individuals possessed vastly different normative understandings of masculine versus feminine tasks, it seems safe to assume that the majority of this study's participants would consider mechanical ability an aptitude descriptive of the masculine gender role and could therefore satisfy this component of Nash's (1979) gender role hypothesis.

On the other hand, there is some question as to whether individuals are capable of meeting the first premise implied by Nash's (1979) theory. As Hamilton (1995) points out, the gender role hypothesis rests on the assumption that individuals are fairly well gender-typed and would thus exhibit a relatively large difference between their masculinity and femininity scores on a measure of gender role identification. By extension, this proposition supposes that individuals are clearly in touch with cultural definitions of masculine and feminine behavior and that they use such delineations to evaluate their self-concept and make judgments concerning the appropriateness of their

behaviors (Ritter, 2004). However, closer empirical examination of this contention (e.g., Twenge, 1997; Holt & Ellis, 1998; Duehr & Bono, 2006) has shown that people no longer appear to identify their personal gender role orientation distinctly enough to achieve the "boost" to cognitive performance gained from a compatible gender role-cognitive task match. Likely owing to significant sociocultural changes in the 30 plus years since Bem (1974) and her contemporaries popularized the study of gender role identification, it has become much more socially acceptable for one to develop and express characteristics indicative of both masculinity and femininity (Twenge, 1997; Ritter, 2004). In so doing, individuals have become much less likely to define themselves as masculine *or* feminine, instead developing a self-concept that more substantially blends characteristics of both masculinity *and* femininity.

For example, a meta-analysis conducted by Twenge (1997) examining changes in masculinity/femininity scores on the BSRI and the Personal Attributes Questionnaire (Spence et al., 1975) between 1973 and 1994 found a significant positive relationship between time and androgyny scores, such that more individuals appeared to be identifying themselves as high masculine and high feminine as time progressed. This correlation could partly be explained by the finding that men demonstrated a minor positive increase in femininity scores, whereas women exhibited a dramatic increase in masculinity scores. A longitudinal analysis of gender stereotypes and managerial success by Duehr and Bono (2006) across a comparable time frame supports a similar conclusion, reporting that females have become increasingly more associated with typically masculine managerial qualities.

Though Twenge (1997) references research and theories that point to a number of possible psychological and sociological explanations for this phenomenon, more important to the present discussion is this trend's applicability to the results observed by the current study. As can be seen in Table 10, the mean scores on the BSRI-masculinity (M) *and* -femininity (F) subscales for the sample were almost exactly 5.00, indicating that, on average, participants generally described their self-concept as possessive of characteristics consistent with both gender roles. Further support for this pattern of androgyny can be found by examining the gender role scores separately for men and women (Table 11); although subjects did identify more strongly with their same-sex gender role (i.e., men more strongly masculine, women more strongly feminine), males and females still tended to score above the midpoint on both scales. Taken together then, one explanation for why gender role identification did not appear to influence mechanical ability in the present study could be that individuals did not perceive themselves as clearly "masculine enough" to gain any potential psychological advantage in the stereotypically masculine cognitive domain. Based on this interpretation, these results seem to support the growing notion that the applicability of the gender role hypothesis in present day American culture has been minimized.

Despite the above interpretation and the advocacy it has received from some scholars (e.g., Signorella & Jamison, 1986; Hamilton, 1995; Ritter, 2004), an alternative view also exists that adopts a much less critical approach of Nash's (1979) hypothesis. In this perspective, the validity of the gender role-cognitive ability relationship is not necessarily discredited; instead, the unreliable relationship between gender role and cognitive ability is posited as a result of poor operationalization of the primary gender-

related constructs involved. As described in detail by Choi and Fuqua (2003), the traditional method of measuring gender role identification (i.e., self-report questionnaires that ask a respondent how adequately a particular behavioral characteristic describes himself/herself) does not appear to be an adequate representation of a person's overall gender role identification. Thus, Nash's (1979) hypothesis may still be correct, but the current tools of measurement do not allow us to observe its proposed relationship.

As support of this claim, Choi and Fuqua (2003) summarized the findings of 23 studies which examined the structure of the BSRI, thus allowing them to describe the most commonly observed factor analytic pattern to emerge across multiple measurement instances and samples. Contrary to a simple two-factor (masculinity-femininity) structure that might be expected and desired, the authors found that anywhere from two to eleven factors were reported across the various studies. Furthermore, the item loadings for each factor were not always consistent, varying somewhat based on the sample used (college vs. non-college populations) and geographic location, among other things. Nevertheless, the most common factor structure found consisted of four factors: a single factor composed only of femininity items, two to three factors composed of masculinity items, and one bipolar factor (commonly labeled a "Sex" factor) on which the two items "masculine" and "feminine" loaded.

In postulating why this factor structure was so commonly observed, Choi and Fuqua (2003) suggest that perhaps the subscales of the BSRI are only capturing a very limited range of the content domain that constitutes "masculinity" and "femininity." For example, the most reliable "femininity" factor to emerge consists of only 10 out of the possible 20 items (affectionate, compassionate, eager to soothe hurt feelings, gentle,

loves children, sensitive to the needs of others, sympathetic, tender, understanding and warm); the remaining items appear to spread out across a number of factors and generate mostly weak loadings. Echoing the sentiments of other recent researchers as well, they posit that this 10-item factor could be more accurately described as an expressiveness/communality factor that occupies only a small portion of the femininity construct space. In this sense, they argue that the BSRI-F in its current form does not accurately measure the construct of femininity as a whole. The same contention is proffered against the BSRI-M, arguing that the two to three most commonly observed factors reflect instrumentality and autonomy rather than the whole of "masculinity."

As was described previously in the Results section, the present study was not concerned with examining the validity of the BSRI as a measure of masculinity and femininity per se and thus the decision to investigate the factor structure of the measure and subsequently create/modify different subscales was not pursued prior to hypothesis testing. However, for the purposes of this discussion, a post hoc analysis was conducted to analyze the factor structure of the BSRI using this study's sample by subjecting the twenty items from both the BSRI-M and BSRI-F to a principal factor analysis followed by a varimax rotation (Table 21). In large part, the results of this investigation revealed a pattern of item loadings remarkably similar to that summarized by Choi and Fuqua (2003). Choi and Fuqua (2003) indicate that the single femininity factor most consistently reported in the literature was composed of 10 items (listed in the preceding paragraph) from the BSRI-F subscale; these exact same items once again loaded on a common factor (Factor I) in the present sample, though here an additional item ("Loyal") also emerged. Additionally, the bipolar Sex factor (Factor IV) was also replicated in the

Table 21

Rotated (Varimax) Factor Loading Matrix for Items from the BSRI

Items	Factor I	Factor II	Factor III	Factor IV	Factor V
Affectionate (F)	.75				
Compassionate (F)	.86				
Eager to soothe hurt feelings (F)	.65				
Gentle (F)	.68				
Loves children (F)	.51				
Loyal (F)	.42				
Sensitive to the needs of others (F)	.79				
Sympathetic (F)	.74				
Tender (F)	.66				
Understanding (F)	.62				
Warm (F)	.70				
Ambitious (M)		.49			
Independent (M)		.74			
Individualistic (M)		.63			
Self-reliant (M)		.65			
Self-sufficient (M)		.76			
Willing to take a stand (M)		.52			
Willing to take risks (M)		.51			
Aggressive (M)			.65		
Dominant (M)			.67		
Forceful (M)			.71		
Feminine (F)				.88	
Masculine (M)				-.84	
Acts as a leader (M)					.79
Has leadership abilities (M)					.61
Eigenvalue	8.32	5.44	2.42	2.03	1.69
% of variance	20.8%	13.6%	6.1%	5.1%	4.2%

Note. The letter in parentheses following the item indicates whether it came from the masculine subscale (M) or the feminine (F) subscale of the BSRI. Factors comprised of items with an absolute loading less than .4 are not displayed to ease interpretation.

current sample. The three remaining factors consisted of seven, three, and two items, respectively, from the BSRI-M subscale. These factors did not reproduce a pattern of

item loadings similar to that reported in Choi and Fuqua (2003), though Factor II could be described as a combination of the previously observed autonomy and instrumentality factors. Items loading on Factor III appear to represent the "machismo/alpha male" mentality commonly associated with highly masculine individuals, whereas Factor V is clearly related to leadership qualities.

Given the above, what do the findings from Choi and Fuqua (2003) and those presented in Table 21 mean in relation to this study's results and the future of gender role research? To begin, one obvious implication concerns the use of the BSRI as an accurate measure of gender role identification. In a lengthy discussion of the ten greatest problems facing the measurement of masculinity and femininity, Beere (1990) points out that issues of content validity and multidimensionality in currently developed measures of gender role identification continue to be two of the most prolific challenges facing gender role researchers. The BSRI is clearly no exception to this criticism; to assume that 20 behavioral characteristics are sufficient to describe a construct as complex as masculinity or femininity, especially considering their dependence on often changing and ambiguous cultural definitions, seems a stretch. What's more, even with the limited number of items currently in use, multiple factors still managed to emerge from the purportedly unidimensional subscales (Bem, 1974). Thus as it currently stands, the present research supports the notion that the BSRI does not appear to be as valid a measure of masculinity and femininity as it was originally intended.

On the other hand, though, these findings could also be taken to suggest that masculinity and femininity should be treated as higher order latent variables that are not directly measurable but which could be estimated from more proximal latent subfactors

using structural equation modeling techniques. To the best of this author's knowledge, such a consideration has not been explicitly examined or considered in the research literature; nevertheless, given the results presented above, it appears a plausible alternative. In this sense, one could argue that the BSRI is useful for capturing a select few lower order factors (i.e., expressiveness, instrumentality, etc.) that describe a portion of the unique construct spaces defined by femininity and masculinity. If this were the case, one could then interpret the non-significant relationships observed in Hypothesis 5 and 6a-6d by stating that the specific *facets* of masculinity and femininity measured by the current study did not appear to influence mechanical ability.

The above conclusion does not necessarily negate the predictions of the gender role hypothesis proposed by Nash (1979), but instead qualifies it by stating that only certain aspects of one's gender role orientation provide a "boost" when performing a gender stereotyped cognitive task. To identify and validate the notion of multiple lower order masculinity and femininity factors would require extensive efforts directed at item development in order to accurately sample the content domain of each gender role construct, followed by repeated factor analyses across multiple samples. However, if such facets could be reliably identified, the boon to future research could be substantial—much in the same way the creation of the "Big 5" reenergized personality research, the same could be achieved in the field of gender role research. Through continued research on and development of measures for lower order masculinity and femininity facets, a new and better understanding of gender role orientation might be achieved that would be much better suited for examining specific gender hypotheses that are currently not adequately explained.

An additional brief note of interest concerns the bipolar Sex factor identified by Choi and Fuqua (2003) and reproduced with this study's sample. Although Constantinople (1973) popularized the notion that masculinity is not simply the opposite of femininity (and vice versa) and thus gender role identification should not be placed on a continuous unidimensional spectrum, the observed Sex factor suggests that individuals do in fact consider these two descriptors in rather mutually exclusive terms—more often than not, respondents believed either masculinity or femininity was the better descriptor of their self-concept, but not both. However, as previous research has shown (e.g., Twenge, 1997) and the current study replicated, individuals are quite likely to endorse characteristics from both the BSRI-M and BSRI-F and therefore be classified as high masculine/high feminine.

How, then, can these seemingly contradictory findings be reconciled? While a definitive answer can not be derived on the basis of the present results, the argument could be made that the Sex factor is merely an artifact of language and represents respondents' conscious effort to avoid the cognitive dissonance that could occur from positively endorsing perceived antonyms in a measure. For example, although not presented in Table 21 because of their small factor loadings, the masculine items "analytical" and "athletic" both grouped on a single bipolar factor as well, suggesting that respondents tended to think of these as opposing descriptors of their self-concept. Given popularly portrayed stereotypes and images of the "dumb jock" and the "frail bookworm," it does not require much imagination to guess why individuals might consider these two terms incompatible and thus endorse only one or the other as part of their own self-concept.

In similar fashion, perhaps individuals felt uncomfortable stating they were simultaneously masculine and feminine knowing that the two terms are often associated with seemingly opposing personal characteristics. These aspects of social desirability are definite issues of concern when using self-report, personality-type questionnaires— though little is known about how social desirability affects participants' responses on the BSRI. Although the BSRI contains a social desirability subscale, it is unclear how this factor is used (or was intended to be used), and a number of the items are likely not as "neutral" as they were intended (see Appendix D). However, both of these conjectures are speculation at best; a more explicit examination of the Sex factor and social desirability in the BSRI is needed in future research to better understand how these issues affect the construct validity of the measure.

The second and final gender-related characteristic examined in this study's model was gender stereotype endorsement. Stereotype endorsement had previously been identified as a significant moderator of relationships between various individual difference variables and performance-related outcomes in other works (e.g., Levy et al., 1998; Blanton et al., 2002; Schmader et. al., 2004), and thus was included as a moderator in the present analyses as well. However, the endorsement of male or female stereotypes did not influence mechanical self-efficacy (Hypothesis 4) or performance on the mechanical comprehension test (Hypothesis 5) in the manner predicted, making an interpretation of this variable as a useful predictor of gender differences difficult. As was the case with gender role identification, this issue of non-significance again begs the question: Why were similar significant relationships not observed in the present experiment though strongly suggested by previous research?

One potential explanation centers on the operationalization of gender stereotype endorsement in general, and the measurement strategy employed in the current study in particular. Although this study attempted to use a nearly identical measure of stereotype endorsement as that implemented by previous researchers, the issue of measurement contamination may have been a legitimate concern here and one that could have influenced the outcomes observed in this study. Similar to past research, the present experiment defined stereotype endorsement as the degree to which an individual believes a particular characteristic about males or females is true. To operationalize this definition, items in the measure were constructed so as to read in the following manner: "Men have better _____ skills than women" (or "Women have better _____ skills than men"). In some respects, indicating whether this item truthfully reflects a characteristic of men (or women, as the case may be) does appear to satisfy the intended nature of the measure; however, this structure also adds an unnecessary comparative component by forcing a frame of reference on respondents. By definition, stereotypes are merely characteristics mentally associated with a social category (Stangor & Lange, 1994); they suggest nothing about whether the same characteristics can or can not apply to a greater/lesser degree for other social categories. In this sense then, stereotype endorsement may *imply* a relative judgment, but it does not explicitly *require* one. For example, the same question from above might have been re-written as: "On average, men are very good at things that require _____ skills." Responding to the truthfulness of this item still provides an indication of one's belief in the validity of the particular stereotype at hand, but does so without imposing an explicit comparison between males and females.

The reason this comparative framework is potentially worrisome is because it introduces the very real possibility of a confounding effect based on individuals' desire to promote positive views of their social group, which in this case refers to their sex. Tajfel and Turner's (1986) work in social identity theory posits that feeling as though one "belongs" to a particular social network or group of like-minded people provides one of the primary sources of reinforcement for individuals. Similar to in-group/out-group theories of social comparison, people attribute the positive characteristics of their social groups to their own self-concepts and thus benefit from identifying with a group that possesses desirable qualities. In turn, people should be motivated to advocate that the groups they belong to are generally "superior" (or, at the very least, are no worse off) than other groups as a means of improving, maintaining or protecting their own positive sense of self. In the present study, the only groups made explicitly salient to respondents were males and females. Thus, asking women to directly endorse a stereotype that males are better at something than females is, according to social identity theory, akin to asking them to freely admit their social group is of a lower status (the opposite would be true for males as well).

Based on this logic, one would expect men and women to most strongly endorse items that reflect positively on their own sex. As all stereotype endorsement items in the present study were worded in the affirmative (i.e., Men/women *better than* women/men at _____), this response pattern would be reflected in the data as females reporting higher ratings on the female stereotype endorsement scale than males (thus more strongly endorsing pro-female stereotypes than males) and males reporting higher ratings on the male stereotype endorsement scale than females (thus more strongly endorsing pro-male

stereotypes than females). As can be seen in Table 11, this is precisely the response pattern observed in the present study. On the female stereotype endorsement scale, women tended to score $d = .61$ standard deviations higher than men, while men tended to score $d = .35$ standard deviations higher than women on the male stereotype endorsement scale, both of which indicated significant mean differences. Taken together, these results appear to be strong evidence that respondents in the present sample were answering in a self/group-reinforcing manner predicted by social identity theory (Tajfel & Turner, 1986) rather than objectively evaluating the validity of a particular stereotype.

How, then, might this finding explain the non-significant effects of stereotype endorsement reported here compared to studies in which the variable was a significant predictor of similar cognitive performance outcomes? For one, the majority of studies reviewed that employed similar measures of explicit gender stereotype endorsement were attempting to elicit social comparison or stereotype threat in their samples (Levy et al., 1998; Blanton et al., 2002; Schmader et al., 2004). Thus, any self/group-reinforcing response pattern that the measure engendered was to be expected and actually desired, as it would then be seen as evidence that the experimental manipulation (e.g., making gender salient to participants, drawing attention to one's group versus personal identity, etc.) was successful. Therefore, in these studies, the pro-social group response pattern would not be perceived as confounding.

However, in the present research, social comparison was not directly manipulated nor was it a desired characteristic of the experimental context, and thus the observed gender differences in the stereotype endorsement measures could not be attributed to controlled variation in the study. As a result, the non-random, non-normal variance

observed in both endorsement measures likely violates the assumption of residual independence; in other words, there is strong reason to believe that being male or female was significantly related to the responses individuals made to both the male and female stereotype endorsement measures (Cohen, Cohen, West & Aiken, 2003). Although nonindependence of the residuals does not affect estimates of regression coefficients, it does affect calculation of the standard errors, which in turn increases the potential for Type II errors. In sum, the fact that gender was indicative of responses to the stereotype endorsement measures likely lessened the probability that any relationships involving the construct would reach statistical significance (for example, the near-significant interaction observed between gender, femininity and female stereotype endorsement in Hypothesis 5, see Figure 8).

Despite the issues associated with poor operationalization and the resultant confounding effect caused by the self-/group-reinforcing response pattern, the present experiment did identify an interesting interactive effect between male and female stereotype endorsement across gender on performance on the mechanical comprehension test (see *Male-Female Stereotype Endorsement* in the Additional Analyses section). To summarize briefly, a main effect of female stereotype endorsement was found for females such that higher levels of endorsement predicted poorer test performance. However, for males, a significant two-way interaction emerged between male and female stereotype endorsement such that men who strongly endorsed both male and female stereotypes performed significantly worse on the test compared to men who more strongly endorsed male stereotypes as opposed to female stereotypes.

Two tentative insights can be drawn on the basis of these findings. The first concerns the inverse relationship observed between female stereotype endorsement and BMCT test performance. As the analysis of Hypothesis 5 demonstrated, the negative main effect of female stereotype endorsement on mechanical ability either approached or attained statistical significance in both regression equations for which it was estimated (see Table 15), lending a moderate degree of stability to the result reported above. But why would one's belief in the validity of stereotypes about women influence performance on a test of mechanical ability? In the introductory sections in which the justification for the proposed gender differences model was presented, the argument for including gender stereotype endorsement as a meaningful variable was based on the belief that it produced an effect similar to stereotype threat (e.g., Steele & Aronson, 1995; Schmader, 2002; Schmader et al., 2004). Given the preceding discussion on social identity theory (Tajfel & Turner, 1986), the endorsement of female stereotypes might also be perceived as an indirect measure of one's identification with the female *sex*. Together, these two interpretations could be taken as evidence to suggest that individuals who strongly endorsed female stereotypes were more fully aware of their "femaleness," and thereby experience a stereotype threat-like effect in which this female quality was not seen as conducive to performing well on a cognitive ability test typically associated with male superiority. While this finding has been somewhat elusive in previous research, the notion that gender identification can play a significant role in the saliency and potency of stereotype threat effects on test performance has been empirically demonstrated (Schmader, 2002).

Without question, this argument rests on the debatable assumption that the endorsement of male/female stereotypes is significantly correlated with one's level of gender identification with the male/female sex. It would certainly seem plausible to imagine a scenario in which a highly female identified individual would not endorse positive stereotypes regarding the female gender. Nevertheless, the implications from Tajfel and Turner's (1986) social identity theory would seem to favor the prediction that a relatively strong relationship exists between gender stereotype endorsement and gender identification, *especially if the stereotypes being endorsed are largely positive in nature* (as in the present study). Again, if individuals seek reinforcement from their group identification as the theory suggests, they should be more likely to endorse positive stereotypes about their group to reinforce their personal belief structures. Unfortunately the present research is unable to test this supposition as gender identification was not included among the study's measured variables; future research would be needed to clarify whether this relationship can be empirically reproduced.

Although the above claim is somewhat speculative, what makes it a potentially intriguing interpretation is that whereas stereotype threat is a function of the experimental context and caused by a purposeful manipulation from the researcher, the effects of stereotype endorsement come directly from the implicit cognitions or attitudes of an individual (Nosek et al., 2002). By this definition, then, one might alternatively perceive stereotype endorsement as a "self-constructed" stereotype threat effect. In other words, because stereotype endorsement is proposed to be a direct outcome of a self-evaluative process in which respondents tap into past experiences, cultural norms, personal beliefs, etc. in order to assess the validity of a given stereotype (Schmader et al., 2004), the

individual could be viewed as the primary cause for creating the threatening situation. From a practical standpoint, such a possibility has very real implications for the strategies one might pursue to reduce gender differences in test performance and the subsequent development of more "gender-friendly" tests. Although stereotype threat research would suggest that the removal of any material that could potentially arouse awareness of gender stereotypes or one's gender identity from a testing situation would prevent/alleviate large sex differences in test performance, the above interpretation implies that people carry these pieces of information around with them all the time and could therefore be aware of them regardless of the contextual information present in the test environment. Thus, a better way to minimize gender differences in such situations might be teaching individuals test taking strategies that reduce the saliency/importance of one's gender identity in relation to test performance and thus prevent invoking possibly debilitating thought patterns. However, as the present research does not specifically test nor do its results specifically support such a notion, future research would be needed to examine the validity of this claim.

The final point of interest in relation to the three-way interaction found between gender, female stereotype endorsement and male stereotype endorsement is closely related to the discussion above, though it concerns the significant two-way interaction observed for male participants between female and male stereotype endorsement. Again, the shape of this interaction was such that although female stereotype endorsement had no effect on test performance for low male endorsing men, high male endorsing/high female endorsing individuals did significantly worse on the mechanical ability test than high male endorsing/low female endorsing individuals. An interpretation of this finding

does not follow clearly from the previous conceptualization of stereotype endorsement as an indirect indicator of gender identification (Tajfel & Turner, 1986) and a self-imposed stereotype threat effect. Males who strongly endorsed female stereotypes (indicating an identification with the female sex) only appeared to do poorly on the mechanical ability test when their level of male stereotype endorsement was high as well (indicating some identification with the male sex as well); when female endorsement was high, but male stereotype endorsement was low, performance on the test was not attenuated.

Unfortunately, the broader research literature on stereotypes and social identity offers little help in suggesting why these results were observed in the present study. Perhaps one explanation, however, could be surmised by considering the possible situational demands associated with male performance on a test of mechanical ability. One intriguing line of stereotype threat research suggests that males perform better on a stereotypically male-advantaged/female-disadvantaged cognitive ability test (such as the BMCT) through the process of stereotype lift (Walton & Cohen, 2003). In brief, stereotype lift proposes that a performance boost may be experienced by members of a particular social in-group when they are made aware of a negative stereotype about the abilities of an opposing out-group which can subsequently be used as a base of comparison. Similar to the rationale behind a self-constructed stereotype threat effect then, a similar self-constructed stereotype lift effect might exist in which the information required to identify one's relative membership status and their comparative standing versus the out-group is already possessed by an individual.

If this were the case, stereotype lift would be demonstrated in the present sample if high male endorsing/low female endorsing men (i.e., "male affiliated" men) performed

111

better than a low male endorsing/low female endorsing or low male endorsing/high female endorsing male (i.e., not "male affiliated" men). Although this precise finding was not statistically significant in the present dataset, there did appear to be at least some evidence to suggest such a phenomenon could be taking place. A close examination of Figure 7 does show a slightly positive slope for low female stereotype endorsing men as they more strongly endorse male stereotypes. Again, the results from this study do not provide conclusive enough evidence to support this conjecture, but perhaps with a larger sample of males (n = 99 men in the current sample) a more pronounced and statistically significant difference would emerge.

How, though, might the rationale behind stereotype lift explain the poor test performance of high male endorsing/high female endorsing men in relation to other males in the sample? The argument is once more speculative in nature, but social identity theory offers one possible solution to this quandary (Tajfel & Turner, 1986). Although this portion of the male sample tended to report in a self-/group-reinforcing pattern that indicated their "maleness," they also were also equally aware of their "femaleness." This being the case, it could be argued that such high male endorsing/high female endorsing men would experience some manner of cognitive dissonance if they were to benefit from their "maleness" by denigrating their "femaleness." Their identification with the female sex and the associated dissonance could have thus resulted in the observed pattern of diminished performance. Although circuitous support at best, it is interesting to note that high male endorsing/high female endorsing men achieved performance levels on the BMCT (M = 40.0) comparable to those reached by the top-performing females (M = 41.5). In this sense, one could argue that high male endorsing/high female endorsing

males still performed relatively well in relation to the majority of females (thus satisfying the quality of their "maleness" that implies they should do well on a test of mechanical ability), but without the advantage of stereotype lift that would have been gained by debasing the strongly felt "female" portion of their self-concept (thus maintaining a positive perception of their "femaleness").

In summary, the gender-related characteristics of the proposed gender differences model did not generally influence mechanical ability test performance or mechanical self-efficacy in the manner in which they were hypothesized (Hypothesis 4, 5, and 6a-6d). Upon closer examination though, a number of intriguing alternative explanations and implications were observed that suggest potentially new and as yet unexplored avenues of research. However, a cautionary word of warning is warranted at this point against placing too much credence in the theoretical interpretations proffered above without further and more deliberate empirical investigation. Nevertheless, continued exploration of gender-related variables such as gender role identification and gender stereotype endorsement would undoubtedly improve researchers' and practitioners' understanding of what it "means" to be men or women in relation to describing male-female differences, and ultimately allow for a better appreciation of what and how other gender associated influences affect cognitive ability test performance.

Contextually-related characteristics. With only the one exception noted above (Hypothesis 4), the remaining hypotheses examining the contextually-related characteristics of the gender differences model (Hypotheses 1, 2, 3a/3b) were well supported. This suggests that mechanical interests, experiences and self-efficacy play integral roles not only as predictors of mechanical ability, but also as meaningful sources

of variation between males and females. As the results of this study appear to support the theoretical rationale and arguments already made concerning the impact of contextually-related characteristics on cognitive ability performance, the present section will instead focus briefly on what the significance of the present findings suggest for research and practice aimed at minimizing gender performance discrepancies in future testing instances.

As predicted by Bandura (1977) and later formulations by Gist and Mitchell (1992), mechanical interests and experiences were strong predictors of an individual's reported level of mechanical self-efficacy. Furthermore, the predictions drawn from Bandura's (1986) social cognitive theory of performance were also supported by the finding that mechanical self-efficacy was predictive of performance on the mechanical ability test. These linked relationships supported the proposed mediation hypothesis, indicating that both mechanical interests and mechanical experiences were predictive of performance on the BMCT, though largely through their positive effects on mechanical self-efficacy. Based on these results, at least two conclusions of practical significance can be drawn. First, the original creators of the BMCT specifically designed the test so as not to favor individuals who had previous training in physics or other related specialties—operationalized here as mechanical experience—but instead to capture one's aptitude for perceiving and identifying physical principles in everyday life occurrences (Bennett, 1969). Although the regression analyses presented for Hypothesis 2 would appear to support this a priori presumption (via the observed non-significant beta coefficient for mechanical experiences), there is strong evidence that the simultaneous inclusion of mechanical interest in the regression equation acted as a suppressor. As

Cohen et al. (2003) suggest, a suppressor effect can often be identified by examining the direction of the relationship in a simple correlation table with the observed beta weights from regression; if a change in sign (+/-) is observed, it is likely that the effects of the variable are being masked by another included predictor. As this was indeed the case found in the current results (cf. Table 10 and 13/18), the observed zero-order correlations should be a better indicator of the actual relationship between mechanical experience and BMCT performance. Thus, the results of this study do not appear to support the presumption that mechanical experiences are not indicative of performance on the BMCT.

Although care should be taken in inferring such strict causal inferences from correlational research, this the predictive relationship between mechanical experience and mechanical ability test scores comes as no great surprise. If one considers specific cognitive ability a form of specialized intelligence (which is not a significant conceptual leap given the theoretical similarity shared between g, s, and measures of intelligence, Spearman, 1927; Vernon, 1950) there is a growing body of research spearheaded by Sternberg (e.g., 1998; 1999; 2001; 2005) and colleagues (e.g., Sternberg, Grigorenko, Ferrari, 2002) that suggests the same processes/techniques used to develop expertise are applicable to developing intelligence. While the literature base on expertise is far too large and extensive to describe in much detail here, empirical studies have generally supported the notion that continued experience and deliberate practice within a given domain greatly facilitates the acquisition of the skills and abilities necessary to reach a high degree of proficiency related to that area of functioning (cf. Ericsson, Charness, Feltovich & Hoffman, 2006). Thus, despite what the original test publishers may have posited, individuals seeking to improve their mechanical comprehension and subsequent

performance on related aptitude tests would seemingly be well served to tackle minor household maintenance issues, tinker with common repair jobs, briefly examine how a particular mechanical object functions, and take advantage of the myriad other similar experience-building opportunities in between. Furthermore, educators and vocational specialists seeking to help students or clients excel on mechanical ability tests might consider adding more practical demonstrations and learning tools to their curriculums to encourage expertise development and familiarity with the common physical principles included in the test materials.

The second conclusion of significance that can be drawn from the present results concerns the specific role mechanical interests and mechanical self-efficacy play in relation to mechanical comprehension test performance. Although self-efficacy did mediate the relationship between interests and experiences with performance, the results suggested the relationship was only one of *partial* mediation—meaning both mechanical background variables were still somewhat predictive of scores on the BMCT above and beyond one's reported level of mechanical self-efficacy. This implies two additional points of leverage for individuals or educators seeking to improve performance on the mechanical ability test. First, because higher self-efficacy scores were correlated reasonably strongly with scores on the BMCT, improving an individual's confidence in his/her ability to perform mechanical activities may result in better overall performance. As presented previously, Bandura (1977; 1986) and Gist and Mitchell (1992) present a useful theoretical framework from which strategies for improving self-efficacy may be selected and pursued further.

116

It should be noted that there is emerging empirical evidence to suggest there are limits on the degree to which an individual's self-efficacy can be meaningfully enhanced and whether self-efficacy is even predictive of improved performance across multiple measurement trials (e.g., Vancouver, Thompson, & Williams, 2001; Vancouver, Thompson, & Tischner, 2002; Vancouver & Kendall, 2006). While the present study did not adopt a within-person design and thus can not directly support or deny these arguments, these results do not seem particularly troublesome given the typical administration of ability tests such as mechanical comprehension. In large part, the BMCT is used as a one-time, applicant or employee screening tool for a variety of purposes (Super & Crites, 1962); as such, it would be somewhat unusual for an individual to take the test more than once in close succession. Thus, the initial, "between-person" level of mechanical self-efficacy one possesses is likely still an important influencer of subsequent performance on a mechanical comprehension test, and any effort to improve those beliefs prior to taking the test should still be beneficial.

The final point of leverage for improving BMCT performance implied by the partially mediated mechanical interests-mechanical comprehension relationship is not a particularly insightful one, but important nonetheless: while acquiring more domain-related interests should correlate with improved self-efficacy, it also appears to influence performance above and beyond the effects of these efficacy beliefs. In this sense, creating genuine interest in mechanically-related subject matter for an individual clearly indicates a highly effective source for improving mechanical comprehension scores. Unfortunately, according to many published accounts, interests are not a particularly malleable individual difference. For example, Hall (2002) reports that by age 21 a

person's interests largely stabilize; in fact, it is not uncommon to find test-retest correlations as high as .70 on many interest inventories, even over a period of 20 years. Nevertheless, to the extent one could generate interest in mechanical activities and the like (see Hidi & Baird, 1988, for a brief review of popular "interest development" strategies and an example of one such attempt at generating interest in expository reading), the benefits to performance on a mechanical ability test such as the BMCT appear to be great.

Capturing Male-Female Differences: An Investment Well-Spent

Consistent with previous research (e.g., Bennett & Cruikshank, 1942; Bennett, 1969; Maccoby & Jacklin, 1974; Anastasi, 1981, 1988; Antill & Cunningham, 1982; Feingold, 1988; Stanley et al., 1992; Stumpf, 1995; Halpern, 2000), significant and large performance differences between males and females were observed in mechanical comprehension. Males again outperformed females on the administered mechanical ability test, on average scoring more than one standard deviation better than their female counterparts. However and perhaps more interesting from a conceptual viewpoint, meaningful variation across sex was also found for *all* of the measured psychological variables captured in the present experiment (see Table 11). This observation is particularly meaningful given the oft-adopted approach in the literature whereby gender differences in cognitive ability are so often simply summarized in the traditional, categorical "participant sex" variable with no consideration as to why this might be the case.

At the risk of crossing from theoretical discourse into facetious commentary, what is so unfortunate about this empirical practice is that despite every breakthrough in

bioengineering, genetics and surgical procedure, biological sex is one of the few things that a person can never fully rid themselves of or change for good. Thus, what good do psychologists, as scientists and practitioners of human behavior, achieve by continuously identifying those areas of functioning that exhibit male or female superiority without further examining *what* makes a man or a woman better/worse at something? Armed with that knowledge, one is simply able to make an observational statement. As the model employed by the current study demonstrates, though, there appears to be relatively strong evidence to suggest that the noted sex differences in mechanical ability test performance can be attributed to differences in interests, experiences, and self-efficacy beliefs, which are subsequently influenced by gender role orientation and identification. Armed with *this* knowledge, future research could attempt to devise experiments that manipulate or otherwise alter these variables in some meaningful fashion to determine whether strategies or methodologies can be constructed that effectively and substantially reduce mechanical ability differences. In this manner, observation moves into explanation, explanation into action, and action hopefully into useful implementation, which should be the ultimate goal for any stream of scientific research.

If empirical inquiry does not employ such specific variables or make attempts to investigate such questions with theoretically sound models, gender differences research can easily digress into providing trivial "because he's a man" or "because she's a woman" explanations for male-female discrepancies. Clearly the functional model of gender differences around which the present research was organized did not completely describe why males outperform females on tests of mechanical ability. However, that was not the expectation for the model, nor should it have been. As the brief review of

biological explanations for differences in mechanical ability pointed out, there is definite

reason to believe that men and women *are* hard-wired to be more or less capable of

performing across various domains. What this model does represent, though, is an effort

to describe male-female differences in mechanical ability with individual differences in

psychological functioning rather than "participant sex"—with the hope being that future

research would benefit from the identification of potentially more manipulable variables.

Limitations

Like any empirical investigation, there are several limitations that should be taken

into account when considering the results of the present study. The first such restriction,

which has already been alluded to in earlier discussions, concerns the correlational nature

of the research design. Importantly, as there were no direct manipulations applied to the

testing scenario or to the study's participants, there is no evidence to support causal order

in the proposed gender differences model. Although theoretical rationale and previous

research in the broader literature were used to inform the implied direction of influence

between study variables, without experimental control the results of this study can only

hint at causation, not confirm it.

An additional possible limitation concerns the use of college student as the

primary sample for this study. Aside from potential concerns of realism and external

validity, the research presented earlier on changing perceptions of gender appropriate

roles (e.g., Twenge, 1997; Holt & Ellis, 1998; Ritter, 2004; Duehr & Bono, 2006) implies

potential generational differences in the manner by which individuals would respond to

the gender-related measures of this study. For example, Table 11 indicates that the

degree respondents tended to not strongly endorse stereotypes about women or men,

regardless of their personal sex. However, this may be a symptom of the increasing "androgynization" of more recent generations—wherein both men and women no longer see great differences between the sexes (Twenge, 1997)—though a higher degree of baseline endorsement might be expected in the older, more "traditional" working population. Although it is unclear whether this conjecture has any real basis without further sampling from a more diverse population, it is noted as a limitation of the present study.

The remaining two limitations deal with the measurement of the study's primary constructs. As suggested in the test manual that accompanied the BMCT (Bennett, 1969), the present experiment administered the mechanical ability test as a timed, 30-minute exam. This protocol was followed to allow for an easy comparison of the overall performance of the present sample with that of the normative performance standards gathered for the BMCT. In hindsight, though, this was likely an unnecessary aspect of the testing situation given the purpose of the research at hand. The time limit, while possibly increasing the realism of the testing context, could have been problematic had the time pressure somehow differentially affected male and female test takers (or those high or low on one of the individual difference variables of interest) by forcing those individuals to rush through the test, change their test-taking strategy, encourage guessing, etc. While there is no way to know whether this was the case in the present study, it is an easily remedied problem. Thus in future research in which the purpose of administering a cognitive ability test is simply to determine an individual's absolute level of aptitude, any time restriction unrelated to the desired research question should be removed so as to

prevent the possibility of introducing an unnecessary confound into the testing environment.

Lastly, because all data used in this study were gathered with multiple-choice, self-report questionnaires and testing instruments, mono-method biases or common method variance may have been a concern as well (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). However, as Spector (2006) argues, common method variance is not invariably applicable simply because a single method of data collection was implemented over the course of a study. Instead, one is better served by considering the nature of the constructs being measured and any potential alternative, underlying explanations that could have systematically altered participant responding to those measures. As such, given that that the present study employed two classes of relatively similar variables (i.e., gender-related characteristics and contextually-related characteristics) response patterns reflecting social desirability or a consistency motif would have been the most likely causes of common method variance. Both of these factors generally result in inflated correlations among the similar classes of variables, which could potentially bias the results of subsequent hypothesis tests. However, aside from the relatively high correlations observed between mechanical interests, experiences and self-efficacy, the interscale correlations among the measured variables were in the small to moderate range; thus, if common method variance was an issue, it likely did not contribute significantly to the response patterns of participants. Nevertheless, future research should consider using additional data sources (i.e., behavioral data, peer reports, etc.) to avoid any potential problems associated with mono-method biases.

Conclusion

Although the acceptance of women into traditionally male dominated lines of work has advanced considerably since Frances Gage addressed the American Equal Rights Association nearly a century and a half ago (Stanton, Anthony, & Gage, 1882), the shift continues to be a slow process (Blau & Hendricks, 1979; England, 1981). The passage of Equal Employment Opportunity laws and federal contract programs has helped to speed the transition, but a large number of occupations continue to show consistent and significant inequalities in terms of the overall gender distributions of its employees (Beller, 1982). The purpose of the present research was to test a functional model of psychological antecedents proposed to predict the disparity in performance observed between males and females on tests of mechanical comprehension, a factor which may significantly contribute to the noted occupational segregation trend in vocations that require tests of mechanical ability for entry. Although not all the included relationships hypothesized in the model were supported, a number of intriguing results were discovered that offer incremental contributions to the literature's current knowledge base on gender differences in cognitive ability performance. Nonetheless, continued efforts are needed to uncover further categories of individual difference variables that explain why males and females appear to exhibit superiority across various domains of psychological functioning, in the hopes that we may one day fully understand the processes through which such differences arise and live on through generations of human beings.

Demographics/Background

Instructions:

Please respond to the following questions to the best of your ability.

1. What is your age? _____

2. What is your gender?
    A. Male
    B. Female

3. Are you of Hispanic origin?
    A. Yes
    B. No

4. What ethnicity do you consider yourself to be?
    A. American Indian/Alaska Native
    B. Asian
    C. Black or African American
    D. Native Hawaiian and Other Pacific Islander
    E. White
    F. American Indian/Alaska Native *and* White
    G. Asian *and* White
    H. Black or African American *and* White
    I. American Indian/Alaska Native *and* Black or African American

5. What was your mother's primary occupation when you were growing up?

    _____

6. Did your mother work primarily outside of the home?
    A. Yes
    B. No

7. What was your father's primary occupation when you were growing up?

    _____

8. Did your father work primarily outside of the home?
    A. Yes
    B. No

Mechanical Interests, Knowledge and Experiences
(Adapted from Rechenberg, 2000)

Instructions:
Using the 5-point scale presented below, please respond to the following items to the best of your ability.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

Interests

1. I find it rewarding to create or fix something with my hands.

2. I enjoy reading magazines or watching television shows about mechanically related topics (e.g., automobiles, new technologies, gadgets, etc.).

3. I enjoy learning about new techniques for performing mechanical or technical activities (e.g., tips on home improvement/repair, auto maintenance, etc.).

4. When faced with an object that isn't working properly (such as an appliance or a bicycle), I enjoy trying to figure out the causes of the malfunction.

5. I like trying to discover how mechanical devices work (through observation, taking things apart, etc.).

6. I am less interested in knowing how a mechanical device functions as I am in knowing how to use the device. (R)

Knowledge

1. I am often asked to show or explain to others how to operate a piece of mechanical equipment (e.g., run a lawnmower, use a power tool, use a sewing machine, etc.).

2. I have had much INFORMAL training related to mechanical or technical activities (e.g., being taught how to fix a car, leaky faucet, etc. by a relative or friend).

3. I have had much FORMAL training related to mechanical or technical activities (e.g., a school physics course, vocational/on-the-job training, etc.).

4. When attempting to repair or assemble an unfamiliar object, I often can not figure out what to do without referring to an instruction manual or help guide. (R)

5. I can usually figure out what is wrong with an object that is not working correctly.

6. I try to develop strategies or techniques (e.g. trial and error, working backwards, etc.) for approaching mechanical or technical activities.

Experiences

1. When I was growing up, I often helped fix things around the house.

2. I typically make repairs around the house when they are needed by myself rather than ask for help.

3. I have frequently taken part in skilled manual activities (e.g., home improvement, auto repair, sewing projects).

4. I rarely take on do-it-yourself projects that require me to put together or build something (e.g., bicycle, shelf, desk, etc.). (R)

5. I have performed many tasks that require the use of hand tools.

6. I have had many experiences where my mechanical abilities or skills were helpful in fixing a problem (e.g., changing a flat tire, repairing a broken door, etc.).

APPENDIX C

Mechanical Self-Efficacy

Instructions:

Please indicate how confident you are in your ability to successfully complete each of the following tasks along the 7-point scale provided below.

1	2	3	4	5	6	7
Not at all confident		Not very confident		Very confident		Completely confident

1. Figure out how a mechanical item works (e.g., a flashlight, simple engine, etc.) by observing how its internal components operate (gears, belts, switches, etc.).

2. Explain why an object moved or acted a certain way based on what is happening in its surrounding environment.

3. Identify the simple forces of physics that caused an object to move or behave in a particular manner (e.g., momentum, gravity, centripetal force, etc.).

4. Predict what will happen in a situation when a physical element of the environment changes (for example, predicting which way a tube of lipstick will roll along the floor of your car based on the direction you turn).

5. Determine why an object (e.g., bicycle, mechanical clock, etc.) is no longer working correctly.

6. Identify how the basic principles of physics (e.g., friction, pressure) allow a mechanical item to operate.

7. Break down and identify the basic components (i.e. levers, pulleys, screws, etc.) contained within a more complex machine.

8. Recognize basic principles of physics (e.g. forces of motion, gravity, etc.) in everyday life.

Gender Role Identification
Bem Sex Role Inventory (Bem, 1974)

Instructions:

Please read each of the characteristics listed and indicate to what degree each word or phrase accurately describes yourself on the 7-point scale depicted below.

1	2	3	4	5	6	7
Never true of me	Almost never true of me	Sometimes not true of me	Can be not true or true of me	Sometimes true of me	Almost always true of me	Always true of me

Masculine Items	Feminine Items	Neutral Items
49. Acts as a leader	11. Affectionate	51. Adaptable
46. Aggressive	5. Cheerful	36. Conceited
58. Ambitious	50. Childlike	9. Conscientious
22. Analytical	32. Compassionate	60. Conventional
13. Assertive	53. Does not use harsh language	45. Friendly
10. Athletic	35. Eager to soothe hurt feelings	15. Happy
55. Competitive	20. Feminine	3. Helpful
4. Defends own beliefs	14. Easily flattered	48. Inefficient
37. Dominant	59. Gentle	24. Jealous
19. Forceful	47. Gullible	39. Likable
25. Has leadership abilities	56. Loves children	6. Moody
7. Independent	17. Loyal	21. Reliable
52. Individualistic	26. Sensitive to the needs of others	30. Secretive
31. Makes decisions easily	8. Shy	33. Sincere
40. Masculine	38. Soft Spoken	42. Solemn
1. Self-reliant	23. Sympathetic	57. Tactful
34. Self-sufficient	44. Tender	12. Theatrical
16. Strong personality	29. Understanding	27. Truthful
43. Willing to take a stand	41. Warm	18. Unpredictable
28. Willing to take risks	2. Yielding	54. Unsystematic

Note. The number preceding each item reflects the position of each adjective as it actually appears on the survey.

APPENDIX E

Gender Stereotype Endorsement
(Adapted from Levy et al., 1998 and Blanton et al., 2002)

Instructions:
Some stereotypes are true. For example, men are stereotyped as physically stronger than women. And indeed, a number of studies have supported the view of men as physically stronger than women in general.

Listed below are some common gender stereotypes. As you read these, you may feel that some are based on gender differences that really do exist or you may feel that the stereotype has no basis in fact. Read each of the following stereotypes and rate *the degree to which you **personally believe** the stereotype is based on true gender differences* using the 5-point scale below. Please answer each question openly and honestly.

1 Not at all true	2 A grain of truth	3 Moderately true	4 Mostly true	5 Absolutely true

I believe that...

1. Women have better organization skills than men.

2. Men have better spatial skills than women.

3. Women have greater verbal ability than men.

4. Men have better math skills than women.

5. Women have better interpersonal skills than men.

6. Men have better mechanical reasoning skills than women.

7. Women have better clerical skills (i.e., are good at typing quickly and accurately, identifying mistakes in printed documents, etc.) than men.

8. Men have more overall intellectual ability than women.

9. Men have better analytic reasoning skills than women.

10. Women have more creative ability than men.

APPENDIX F

Informed Consent

Please read the information below completely and carefully:

This is a two-part study. For the first portion of this study, we will be asking you to respond to an online questionnaire asking about your interest, knowledge, experiences and confidence level in dealing with mechanically related content, as well as questions about your personality and beliefs. You will also be asked to respond to a small number of demographic items that will help us describe our research sample. We expect that it will take about 30 minutes to complete this first part of the study. For your participation in the first part of the study, you will receive 1 point of subject pool credit.

Following the online portion, you will be asked to come to the testing site and complete a test of mechanical comprehension, which should last approximately 45 minutes. Upon completion of this final half of the study, you will receive an additional 2 points of subject pool credit. Thus, fully participating in both halves of the study will be worth 3 points of subject pool credit (NOTE: You may only participate in the second part of the study if you have previously completed the first part of the study).

There are no foreseeable risks associated with participating in this study. Your name and information will remain confidential. Your privacy will be protected to the maximum extent allowable by law. The data will be saved for at least five years after it is collected and will only be accessible by two faculty researchers and one graduate student. By typing your name below, you indicate that you are free to refuse to participate in this project or any part of this project. You may refuse to participate in certain procedures or answer certain questions. Your participation is completely voluntary. You may choose not to participate at all and may discontinue your participation at any time without penalty or loss of benefits.

If you have any questions or concerns about your participation in this project, please contact James Grand (by phone: (517) 355-2171, e-mail: grandjam@msu.edu, or by appointment: 348 Psychology Building). If you have questions or concerns regarding your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact - anonymously if you wish - Peter Vasilenko, Ph.D. Chair of the Human Research Protection Program by phone: (517)355-2180, fax: (517)432-4503, email: irb@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824.

Please mark the box that says "I agree to give my consent to participate" if you agree to participate in this study. Mark "I do not want to participate" if you do not agree to participate in this study. If you agree to participate, enter your name below and you will be taken to the survey once this step is completed.

☐ I agree to give my consent to participate
☐ I do not want to participate

First name _____ Middle initial ____ Last name _____

APPENDIX G

Participant Debrief

Participant Feedback/Debriefing

Tests of mechanical ability are typically used as part of a selection process to choose applicants for jobs that are mechanical in nature, such as assembly or maintenance workers. As these tests can be used over a wide range of occupations that call for many different job-specific skills, mechanical ability tests are designed to measure a person's general understanding of how the principles of physics and motion work in our everyday lives.

Historically, males have significantly outperformed females on such tests of mechanical comprehension—however, the reasons for this large and consistent performance gap are unknown. The purpose of the research study you have just participated in is to investigate the possible causes and explanations for why gender differences in mechanical comprehension tests exist. Aside from a person's sex, the variables of interest in this study thought to influence mechanical comprehension include:

1. <u>Mechanical interests, knowledge and experiences</u>
2. <u>Mechanical self-efficacy</u> – one's confidence level in their ability to adequately perform a specific task
3. <u>Gender Role Identification</u> – the degree to which an individual's personality characteristics are typically masculine or feminine
4. <u>Gender Stereotype Endorsement</u> – the degree to which an individual believes and endorses the stereotypical differences between men and women
5. <u>Social Penalties</u> – the degree to which an individual has experienced sanctions or been discouraged from pursuing activities related to mechanical ability

Prior to your participation in this study, neither you nor any of the research volunteers were informed of the gender differences typically found in tests of mechanical comprehension. It was necessary to withhold this information because of research findings which suggest that on tasks in which a cultural stereotype exists (i.e., men have higher mechanical ability than women), any information given to test takers that indicates that one group typically outperforms the other can actually increase the performance gap between the two groups. Therefore, care was taken not to present information on the exact purpose of the research study in order to prevent any undesirable influences on overall test performance.

By better understanding why gender differences in mechanical comprehension testing exist, it may be possible to develop more fair and accurate methods of testing to ensure that any individual, regardless of sex, has equal opportunity to participate in jobs in which mechanical ability is important. If you have further questions about the purpose of this study, how your information will be used, or any other concerns, please contact:

<u>James Grand</u>
Phone: (517) 355-2171
E-mail: grandjam@msu.edu
Location: 346 Psychology Building, Michigan State University, East Lansing, MI 48824

Median Split Analyses for Hypotheses 6a-6d

In an attempt to more precisely replicate Antill and Cunningham's (1992) findings, a similar median split procedure and statistical analysis used by those researchers was implemented. Table 22 presents the results of this procedure, which is described in detail below.

To begin, a median split technique was used to divide participants in the current database into the necessary gender roles. A random sample of 75 males and 75 females were chosen from the dataset and used to calculate median scores for the masculinity and femininity subscales of the BSRI[6]. As Spence et al. (1975) suggest, high masculinity is classified as any individual scoring above the median on the masculinity subscale and high femininity as any individual above the median on the femininity subscale (and vice versa for low masculinity and low femininity). Based on these distinctions, the median scores were used as the cutoff point to form the four gender role groups—masculine, feminine, androgynous, and undifferentiated—used by Antill and Cunningham (1982). The masculine category was constructed to consist of high-masculine/low-feminine individuals, the feminine category of low-masculine/high-feminine individuals, the androgynous category of high-masculine/high-feminine individuals, and the undifferentiated category of low-masculine/low-feminine individuals.

---

[6] As Antill and Cunningham (1982) point out in Footnote 6 of their publication, equal numbers of males and females were selected so that both sexes would equally contribute to the observed median scores.

Table 22

Mean BMCT Scores Based on Median-Split, BSRI Gender Role Identification Categories (Hypotheses 6a – 6d)

	Category									
	M + A	U + F	t	F + A	U + M	t	M	A	U	F
Males	46.4 (58)	46.1 (41)	.181[ns]	44.7 (29)	46.9 (70)	1.300[ns]	47.6 (37)	44.2 (21)	46.1 (33)	46.0 (8)
Females	38.3 (72)	38.4 (87)	.031[ns]	38.4 (112)	38.3 (47)	.112[ns]	38.9 (20)	38.1 (52)	37.8 (27)	38.6 (60)

[ns]p = n.s.

Note. M = masculine; A = Androgynous; U = Undifferentiated; F = Feminine.
Numbers in parentheses indicate the number of subjects on which the mean next to it is based.
The t column refers to the results of the t-test between the 2 means preceding it.

The second step performed by Antill and Cunningham was to nest males and females within the gender role categories. Additionally, the authors also examined differences in mechanical ability test performance within the masculine gender role by creating separate high-masculine (masculine + androgynous individuals) and low-masculine (undifferentiated + feminine individuals) groups for both males and females. Although the researchers did not examine performance variations between high-feminine (feminine + androgynous) and low-feminine (undifferentiated + masculine) individuals in their original study, these comparison groups are necessary to examine the possibility of a gender by gender role interaction. Thus in the present research, high-feminine and low-feminine groups were formed for both males and females as well (beginning of Table 22).

The final step in the analyses presented by Antill and Cunningham was to analyze the observed mean differences between the comparison groups to determine if they were significant and in the predicted direction. As stated earlier, Antill and Cunningham did not present results that tested for differences among the 2 gender (male, female) x 4 gender role (masculine, feminine, androgynous, undifferentiated) cell means, instead

focusing on within gender role effects. Thus, to obtain tests of the main effects and interaction in the present study, a two-way ANOVA was conducted treating gender and gender role identification as between-subject factors. The results of this analysis indicate that only the main effect for gender was predictive of performance differences in the BMCT ($F(1, 250) = 49.492, p < .001$), with both the main effect of gender role identification ($F(3, 250) = .734, n.s.$) and the interaction effect ($F(3, 250) = .363, n.s.$) failing to reach significance[7].

In addition to examining differences among the gender by gender role subgroups, differences within gender role were also examined. To do so, a series of t-tests were conducted that specifically compared high-masculine males to low-masculine males and high-feminine males to low-feminine males; the same analyses were then conducted for females. However, no significant differences were observed across any of the comparisons indicating that variations in males' or females' reported strength of gender role identification did not affect their scores on the BMCT.

---

[7] It is worth pointing out, however, that if only the *absolute* differences between cell means are examined (as Antill and Cunningham, 1982, did), the current sample does reproduce the pattern of findings predicted by Hypotheses 6a-6d. Significance testing aside, masculine males (47.6) outperformed feminine males (46.0), who outperformed masculine females (38.9), who outperformed feminine females (38.6). As the two-way ANOVA revealed though, these differences were not significant; it is the author's belief that had the same analysis been performed with Antill and Cunningham's data, the same non-significant result would have been found among their observed cell means as well.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Anastasi, A. (1981). Sex differences: Historical perspectives and methodological implications. *Developmental Review, 1,* 187-206.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Antill, J. K., & Cunningham, J. D. (1982). Sex differences in performance on ability tests as a function of masculinity, femininity, and androgyny. *Journal of Personality & Social Psychology, 42*(4), 718-728.

Bandura, A. (1969). Social-learning theory of identificatory processes. In D. A. Goslin (Ed.) *Handbook of socialization theory and research* (pp. 213-262). Chicago: Rand McNally.

Bandura, A. (1977). Self efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

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

Baron, R. M., & Kenny, D. 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.

Beere, C. A. (1990). *Gender roles: A handbook of tests and measures.* Westport, CT: Greenwood Press.

Beller, A. H. (1982). Occupational segregation by sex: Determinants and changes. *The Journal of Human Resources, 17*(3), 371-392.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting & Clinical Psychology, 42*(2), 155-162.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex-typing. *Psychological Review, 88,* 354-364.

Bem, S. L. (1985). Androgyny and gender schema theory: A conceptual and empirical integration. In T. B. Sonderegger (Ed.) *Nebraska Symposium on Motivation: Psychology of Gender.* Lincoln, NB: University of Nebraska Press.

Benett, G. K., & Cruikshank, R. (1942). Sex differences in the understanding of mechanical problems. *Journal of Applied Psychology, 26*(2), 121-127.

Bennett, G. K. (1969). *Manual for the Bennett mechanical comprehension test, Forms S & T.* New York: The Psychological Corporation.

Bennett, G. K. (2006). *Bennett mechanical comprehension test, Form S.* San Antonio, TX: Harcourt Assessment Incorporated.

Bernard, M. E., Boyle, G. J., & Jackling, I. (1990). Sex-role identity and mental ability. *Personality and Individual Differences, 11*, 213-217.

Blanton, H., Christie, C., & Dye, M. (2002). Social identity versus reference frame comparisons: The moderating role of stereotype endorsement. *Journal of Experimental Social Psychology, 38*(3), 253-267.

Blau, F. D., & Hendricks, W. E. (1979). Occupational segregation by sex: Trends and prospects. *The Journal of Human Resources, 14*(2), 197-210.

Brown, S. M. (1979). Male versus female leaders: A comparison of empirical studies. *Sex Roles, 5*(5), 595-611.

Cantoni, L. P. (1955). High school tests and measurements as predictors of occupational status. *Journal of Applied Psychology, 39*(4), 253-255.

Chen, G., Gully, S. M., Whiteman, J. A., & Kilcullen, R. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology, 85*(6), 835-847.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62-83.

Choi, N., & Fuqua, D. R. (2003). The structure of the Bem Sex Role Inventory: A summary of 23 validation studies. *Educational and Psychological Measurement, 63*(5), 872-887.

Cleveland, J., Stockdale, M. S., & Murphy, K. R. (2000). *Women and men in organizations: Sex and gender issues at work.* Mahwah, N. J.: Lawrence Erlbaum Associates.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249-253.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80*(5), 389-407.

Cox, J. W. (1928). *Mechanical aptitude: Its existence, nature and measurement.* London: Methuen & Co. Ltd.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5-18.

Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin, 129*(4), 569-591.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3). 573-598.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*(2), 135-172.

England, P. (1981). Assessing trends in occupational sex segregation, 1900-1976. In I. Berg (Ed.), *Sociological perspectives on labor markets* (pp. 273-295). New York: Academic Press.

Ericsson, K. A., Charness, N., Feltovich, P. J., Hoffman, R. R. (Eds.) (2006). *The Cambridge handbook of expertise and expert performance.* New York: Cambridge University Press.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*(2), 95-103.

Flynn, J. R. (1984). IQ gains and the Binet decrements. *Journal of Educational Measurement, 21*(3), 283-290.

Geschwind, N., & Galaburda, A. M. (1987). *Cerebral lateralization: Biological mechanisms, associations, and pathology.* Cambridge, MA: MIT Press.

Ghiselli, E. E., & Brown, C. W. (1951). Validity of tests for auto mechanics. *Journal of Applied Psychology, 35*(1), 23-24.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *The Academy of Management Review, 17*(2), 183-211.

Guertin, A. S., Guertin, W. H., & Ware, W. B. (1981). Distortion as a function of the number of factors rotated under varying levels of common variance and error. *Educational and Psychological Measurement, 41*(1), 1-9.

Guilford, J. P. (1947). The discovery of aptitude and achievement variables. *Science, 106*(2752), 279-282.

Guilford, J. P. (1948). Factor analysis in a test-development program. *Psychological Review, 55*(2), 79-94.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Printed classification tests* (AAF Aviation Psychology Program, Research Reports, Rep. No. 5). Washington, DC: U.S. Government Printing Office.

Hall, D. T. (2002). *Careers in and out of organizations.* Thousand Oaks, CA: Sage Publications.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: L. Erlbaum Associates.

Hamburg, D. A., & Lunde, D. T. (1966). Sex hormones in the development of sex differences in human behavior. In E. E. Maccoby (Ed.) *The Development of Sex Differences* (pp. 1-24). Stanford, CA: Stanford University Press.

Hamilton, C. J. (1995). Beyond sex differences in visuo-spatial processing: The impact of gender trait possession. *British Journal of Psychology, 86,* 1-20.

Harrell, W., & Faubion, R. (1940). Selection tests for aviation mechanics. *Journal of Consulting Psychology, 4*(3), 104-105.

Harris, J. (1978). Sex differences in spatial ability: Possible environmental, genetic, and neurological factors. In M. Kinsbourne (Ed.), *Asymmetrical function of the brain.* Cambridge, NY: Cambridge University Press.

Harris, J. (1981). Sex-related variations in spatial skill. In L. S. Liben, A. H. Patterson & N. Newcombe (Eds.), *Spatial representation and behavior across the life span: Theory and application* (pp. 83-125). New York: Academic Press.

Hidi, S., & Baird, W. (1988). Strategies for increasing text-based interest and students' recall of expository texts. *Reading Research Quarterly, 24*(4), 465-483.

Holt, C. L., & Ellis, J. B. (1998). Assessing the current validity of the Bem Sex Role Inventory. *Sex Roles, 39,* 929-941.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75,* 334-349.

Huston, A. C. (1983). Sex-typing. In E. M. Hetherington (Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 387-467). New York: Wiley.

Jacklin, C. N., & Reynolds, C. (1993). Gender and childhood socialization. In A. Beall & R. J. Sternberg (Eds). *The Psychology of Gender* (pp. 197-214). New York: Guilford Press.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits-self-esteem, generalized self-efficacy, locus of control, and emotional stability-with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*(1), 80-92.

Kenny, D.A. (2008). Mediation. Retrieved November 27, 2008 from http://davidakenny.net/cm/mediate.htm#BK.

Klare, G. R., Gustafson, L. M., Mabry, J. E., & Shuford, E. H. (1955). The relationship of immediate retention of technical training material to career preferences and aptitudes. *Journal of Educational Psychology, 46*(6), 321-329.

Lane, J., & Lane, A. (2001). Self-efficacy and academic performance. *Social Behavior & Personality: An International Journal, 29*(7), 687-694.

Lepore, L., & Brown, R. (2000). Category and stereotype activation: Is prejudice inevitable? In C. Stangor (Ed.), *Stereotypes and prejudice* (pp. 119-137). Philadelphia, PA: Psychology Press.

Levonian, E., & Comrey, A. L. (1966). Factorial stability as a function of the number of orthogonally-rotated factors. *Behavioral Science, 11*(5), 400-404.

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421-1436.

Lott, B., & Maluso, D. (1993). The social learning of gender. In A. Beall & R. J. Sternberg (Eds). *The Psychology of Gender* (pp. 99-123). New York: Guilford Press.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

MacKinnon, D.P., Fairchild, A.J., Fritz, M.S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

Massa, L. J., Mayer, R. E., Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*(2), 99-111.

Martin, G. C. (1951). Test batteries for trainees in auto mechanics and apparel design. *Journal of Applied Psychology, 35*(1), 20-22.

Mathieu, J. E., Martineau, J. W., & Tannenbaum, S. I. (1993). Individual and situational influences on the development of self-efficacy: Implications for training effectiveness. *Personnel Psychology, 46*(1), 125-147.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113(1)*, 181-190.

Messent, P. R. (1976). Female hormones and behaviour. In B. Lloyd & J. Archer (Eds.) *Exploring Sex Differences* (pp. 185-212). London: Academic Press.

Milton, G. A. (1957). The effects of sex-role identification upon problem-solving skill. *Journal of Abnormal and Social Psychology, 55*(2), 208-212.

Moritz, S. E., Feltz, D. L., Fahrbach, K. R., & Mack, D. E. (2000). The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research Quarterly for Exercise and Sport, 71*(3), 280-294.

Muchinsky, P. M. (2004). Mechanical aptitude and spatial ability testing. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment, Vol. 4: Industrial and organizational assessment* (pp. 21-33). Hoboken, NJ: John Wiley & Sons, Inc.

Nash, S. C. (1975). The relationship among sex-role stereotyping, sex-role preference, and the sex difference in spatial visualization. *Sex Roles, 1*(1), 15-32.

Nash, S. C. (1979). Sex role as a mediator of intellectual functioning. In M. A. Wittig & A. C. Petersen (Eds.), *Sex-related differences in cognitive functioning: Developmental issues* (pp. 263-302). New York: Academic Press.

Newcombe, N. & Dubas, J. S. (1992). A longitudinal study of predictors of spatial ability in adolescent females. *Child Development, 63*, 37-46.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality & Social Psychology, 83*(1), 44-59.

Patterson, C. H. (1956). The prediction of attrition in trade school courses. *Journal of Applied Psychology, 40*(3), 154-158.

Patterson, D. G., Elliot, R. M., Anderson, L. D., Toops, H. A., & Heidbreder, E. (1930). *Minnesota mechanical ability tests*. Minneapolis: University of Minnesota Press.

Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*(5), 792-802.

Piedmont, R. L., & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO Personality Inventory. *Journal of Psychology, 128*, 255-265.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Rechenberg, C. (2000). *Understanding gender differences in mechanical performance*. Unpublished doctoral dissertation, University of Akron, Ohio.

Rogers, L. (1976). Male hormones and behaviour. In B. Lloyd & J. Archer (Eds.) *Exploring Sex Differences* (pp. 157-184). London: Academic Press.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50*(11/12), 835-850.

Schmitt, N., & Stuits, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9*(4), 367-373.

Schwoerer, C. E., May, D. R., Hollensbe, E. C., & Mencl, J. (2005). General and specific self-efficacy in the context of a training intervention to enhance performance expectancy. *Human Resource Development Quarterly, 16*(1), 111-129.

Sherer, M., Maddux, J. E., Mereandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and validation. *Psychological Reports, 51*, 663-671.

Signorella, M. L., & Jamison, W. (1986). Masculinity, femininity, androgyny, and cognitive performance: A meta-analysis. *Psychological Bulletin, 100*(2), 207–228.

Signorella, M. L., & Vegaga, M. E. (1984). A note on gender stereotyping of research topics. *Personality and Social Psychology Bulletin, 10,* 107-109.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: MacMillan.

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9*(2), 221-232.

Spence, J. T., Helmreich, R., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality & Social Psychology, 32*(1), 29-39.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240-261.

Stangor, C. (Ed.). (2000). *Stereotypes and prejudice.* Philadelphia, PA: Psychology Press.

Stangor, C., & Lange, J. (1994). Mental representations of social groups: Advances in conceptualizing stereotypes and stereotyping. *Advances in Experimental Social Psychology, 26,* 357-416.

Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., & Lupkowski, A. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.) *Talent development, Vol. 1: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 42-65). Unionville, NY: Trillium Press.

Stanton, E. C., Anthony, S. B., & Gage, M. J. (Eds.). (1882). *History of woman suffrage.* New York: Fowler & Wells.

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.

Stein, A. H., & Bailey, M. M. (1973). The socialization of achievement orientation in females. *Psychological Bulletin, 80,* 345-366.

Stenquist, J. L. (1923). *Measurements of mechanical ability.* New York: Teachers College, Columbia University.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science, 26*(1/2), 127-140.

Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology, 24*(4), 359-375.

Sternberg, R. J. (2001). Giftedness as developing expertise: A theory of the interface between high abilities and achieved excellence. *High Ability Studies, 12*(2), 159-179.

Sternberg, R. J. (2005). Intelligence, competence, and expertise. In A. J. Elliot & C. S. Dweck (Eds.) *Handbook of competence and motivation* (pp. 15-30). New York: Guilford Publications.

Sternberg, R. J., Grigorenko, E. L., & Ferrari, Michel (2002). Fostering intellectual excellence through developing expertise. In M. Ferrari (Ed.) *The pursuit of excellence through education* (pp. 57-83). Mahwah, NJ: Lawrence Erlbaum Associates.

Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology, 84,* 959-968.

Stumpf, H. (1995). Gender differences in performance on tests of cognitive abilities: Experimental design issues and empirical results. *Learning and Individual Differences, 7*(4), 275-287.

Super, D. E., & Crites, J. O. (1962). *Appraising vocational fitness by means of psychological tests* (Rev. ed.). New York: Harper.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn Effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32*(4), 349-362.

Tajfel, H., & Turner, J. C. (1986). Social identify theory of intergroup behavior. In W. Austin & S. Worchel (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 7-24). Chicago: Nelson-Hall.

Teasdale, T. W., & Owen, D. R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence, 28*(2), 115-120.

Thurstone, L. L. (1948). Psychological implications of factor analysis. *American Psychologist, 3,* 402-408.

Tiffin, J., Knight, F., & Asher, E. (1946). *The psychology of normal people.* Boston: D.C. Heath and Company.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34*(4), 414-424.

Tipton, R. M., & Worthington, E. (1984). The measurement of generalized self-efficacy: A study of construct validity. *Journal of Personality Assessment, 48*, 545-548.

Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles, 36*(5/6), 305-325.

van den Berg, P. T., & Feij, J. A. (2003). Complex relationships among personality traits, job characteristics, and work behaviors. *International Journal of Selection and Assessment, 11*(4), 326-339.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological-Review, 113*(4), 842-861.

van Engen, M. L., & Willemsen, T. M. (2004). Sex and leadership styles: A meta-analysis of research published in the 1990s. *Psychological Reports, 94*(1), 3-18.

Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91*(5), 1146-1153.

Vancouver, J. B., Thompson, C. M., & Tischner, E. C. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506-516.

Vancouver, J. B., Thompson, C. M., & Williams, C. M. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*(4), 605-620.

Vernon, P. E. (1950). *The structure of human abilities.* London: Methuen.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83*, 586-597.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*(5), 456-467.

Weisen, J. P. (1999). *Technical manual for the Weisen test of mechanical aptitude (WTMA).* Newton, MA: Applied Personnel Research.

Williams, J. E., & Bennett, S. M. (1975). The definition of sex stereotypes via the Adjective Check List. *Sex Roles, 1*(4), 327-337.

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study* (Rev. ed.). Newbury Park, CA: Sage.

Wittenborn, J. R. (1945). Mechanical ability, its nature and measurement, 1: An analysis of the variables employed in the preliminary Minnesota experiment. *Journal of Educational and Psychological Measurement, 5*, 243-262.

Wolff, W. M., & North, A. J. (1951). Selection of municipal firemen. *Journal of Applied Psychology, 35*(1), 25-29.

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin, 128*(5), 699-727.

Woodruff, S. L., & Cashman, J. E. (1993). Task, domain, and general efficacy: A reexamination of the Self-Efficacy Scale. *Psychological Reports, 72*, 422-432.

Yeo, G. B., & Neal, A. (2006). An examination of the dynamic relationship between self-efficacy and performance across levels of analysis and levels of specificity. *Journal of Applied Psychology, 91*(5), 1088-1101.