INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

BEYOND THE BEM SEX-ROLE INVENTORY: A RECONCEPTUALIZATION OF THE CONSTRUCTS OF "MASCULINITY" AND "FEMININITY" AND A REEXAMINATION OF THEIR MEASUREMENT

by

Rose Marie Hoffman

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
1996

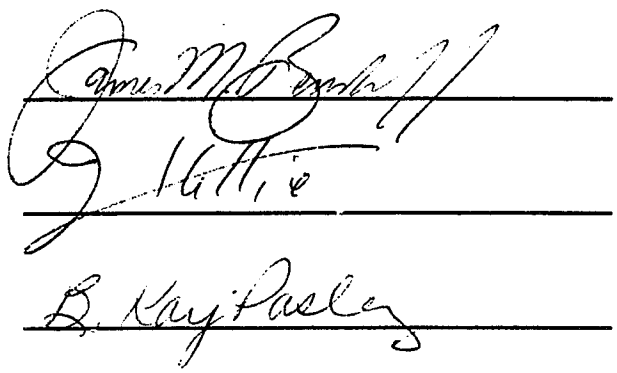
Approved by

L. Dianne Brauer
Dissertation Advisor

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Dissertation Advisor 

Committee Members 

March 18, 1996

Date of Acceptance by Committee

March 18, 1996

Date of Final Oral Examination

HOFFMAN, ROSE MARIE, Ph.D. Beyond the Bem Sex-Role Inventory: A Reconceptualization of the Constructs of "Masculinity" and "Femininity" and a Reexamination of their Measurement. (1996)
Directed by Dr. L. DiAnne Borders, 155 pp.

Current methods of assessing femininity and masculinity reinforce stereotypical gender roles by continuing to label certain characteristics as "feminine" or "masculine." The assessment of femininity and masculinity in terms of stereotypical gender roles has imposed both subtle and severe limitations upon both sexes.

Masculinity and femininity can be reconceptualized in terms of gender identity (Spence, 1984, 1985). Gender self-concept is one component of gender identity, and gender self-confidence can be identified as one aspect of gender self-concept (Lewin, 1984b). As a psychological construct, gender self-confidence is a vehicle by which idiosyncratic definitions of masculinity and femininity can be better understood.

This study was designed to encourage investigation and acceptance of individual meanings that people attach to their perceptions of self as males and females, and to reevaluate stereotypical notions of masculinity and femininity. First, the viability of the Bem Sex-Role Inventory (BSRI) as a research tool was examined by assessing whether its "masculine" and "feminine" items represent current perceptions of masculinity and femininity among college undergraduates. The psychometric development of the BSRI, as well as its theoretical underpinnings, were further examined. Then, based upon its status

as a component of gender self-concept, identified as an aspect of gender identity, gender self-confidence was assessed through the development of the Hoffman Gender Scale (HGS).

Results indicated that current college undergraduates (N = 371) did not perceive BSRI items in gender-linked terms. Analyses of the HGS clearly indicated the existence of two factors that defined gender self-confidence, identified as gender self-definition and gender self-acceptance. Both HGS dimensions were determined to be independent of participants' BSRI classifications.

Utilization of the median-split and hybrid scoring methods across both forms of the BSRI resulted in widely diverse scoring classifications among participants. Implications for past and current use of the BSRI were explored.

Results of the study supported the use of the HGS as a tool to assess gender self-confidence. Implications for subsequent research to further explore the constructs of femininity and masculinity were discussed.

ACKNOWLEDGMENTS

These acknowledgments cannot be free of sentimentality, for then they would not convey the meaning behind them. To the chair of my doctoral committee, Dr. L. DiAnne Borders, I am forever grateful for her constant regard, perception, support, expertise, professionalism, and trust. As a teacher and mentor, she is exemplary. I am indebted to Dr. John Hattie, who strengthened my desire to challenge, shared my love of discovery, and whose boundless energy was a constant source of inspiration. I thank Dr. Kay Pasley and Dr. James Benshoff for their ongoing belief in my work, and for providing me with assistance whenever I asked. The members of my doctoral committee will always hold a special place in my heart.

In a variety of ways too extensive to enumerate, family, friends, colleagues, and others have all contributed to this success. In particular, I must thank Amy, for her dedication and abilities at the computer; Tai, for keeping me in line, and Peaches, whose love and spirit will be with me always.

Finally, one person is as responsible as I am for the completion of this project. To John, who truly made this a joint venture, and who supported, encouraged, loved, entertained, and respected me throughout, I celebrate this accomplishment with you.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

The development and construction of the Bem Sex-Role Inventory (BSRI; Bem, 1974) was based on the conviction that human characteristics (e.g., independent, assertive, gentle, etc.) that can be identified as socially desirable for one of the two sexes should not be used to characterize the other sex. That is, all 20 "feminine" items and all 20 "masculine" items had to be "judged to be significantly more desirable in American society for one sex than for the other" (Bem, 1981a, p. 11) in order to qualify as test items. This line of thinking, as well as the methodology that accompanied it, has been described as unsound by a substantial number of researchers (e.g., Lewin, 1984b; Locksley & Colten, 1979; Myers & Gonda, 1982; Pedhazur & Tetenbaum, 1979; Spence, 1985, 1993). These two areas of concern (i.e., the conceptual basis for the BSRI, and the methodology by which it was developed) are further discussed below.

The conceptual framework underlying the development of the BSRI must be addressed first. According to Spence (1985), "the fundamental logical error that seems to have been made is to assume that an aggregation of statistical facts distinguishing between two groups of individuals, in this instance, men and women, can automatically be combined to arrive at portraits of the typical member of each group" (p. 77). Spence further argued that the lack of strong

correlations among the items within each group (masculine items with other masculine items; feminine items with other feminine items) indicates that few males and females exhibit all or even most of the traits designated "typical" of their gender. The important notion of within group variability as it compares to between group variability remains virtually ignored. Similarly, Lewin (1984b) stated that "this conceptual confusion...led directly to the chief methodological error, the failure to recognize that validating the tests against within-sex behavioral criteria was absolutely critical" (p. 198).

In terms of test construction, the method of item selection for the BSRI has been compared to a method sometimes encountered in achievement testing, "where exclusive reliance on the statistical characteristics of the items for their selection may lead a test constructor to neglect the most important property of the test: its validity" (Pedhazur & Tetenbaum, 1979, p. 998). Bem's (1981a) criterion for an item's inclusion as "masculine" or "feminine" was that it be judged significantly more desirable for a man than for a woman (masculine) or for a woman than for a man (feminine), using independent t-tests. This approach may have resulted in the inclusion of "masculine" and "feminine" items that are not necessarily desirable for one sex over the other, but rather items that are *less undesirable* for one of the sexes, according to Bem's judges (Pedhazur & Tetenbaum, 1979). What often happens, and what seems to have occurred in the construction of the BSRI, is that statistically significant results obscured substantively meaningful findings.

The assumption underlying the development of the BSRI, that "masculine" traits and "feminine" traits are mutually exclusive (i.e., if "independent" is considered masculine it cannot be considered feminine), is an erroneous assumption (Lewin, 1984b), whether the setting is the 1970s or the 1990s. This assumption underlies the conceptual and methodological issues discussed above. An even more fundamental problem with this assumption relates to the question of what is being measured by the BSRI. The constructs under scrutiny here, masculinity and femininity, remain inadequately defined. Although Bem labeled her scales "Masculine" and "Feminine," and claimed interest in examining masculinity and femininity, she seemed to be less than clear about what that meant. In contrast, Spence and Helmreich (1978) argued that their instrument, the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974), although widely acclaimed as the second most popular "MF" test, is basically a measure of instrumentality and expressiveness. Spence and Helmreich (1978) are among many scholars (reviewed in Lippa, 1985) who have argued that the BSRI is basically a measure of instrumentality and expressiveness as well. Unfortunately, the lack of a clear and consistent definition of the masculinity and femininity constructs further belies the use of the BSRI as a measure of masculinity and femininity.

That there are problems regarding the BSRI and its use seems clear. However, in the absence of a more meaningful assessment tool, results of myriad studies using the BSRI are misinterpreted by researchers to confirm

hypotheses that lack adequate theoretical support. False conclusions are drawn using a combination of two fallacious processes. First, men and women are inappropriately defined and labeled in terms of their "masculinity" and "femininity." Second, relationships are suggested to exist between "masculine," "feminine," "androgynous," or "undifferentiated" individuals and various other traits, roles, or behaviors. At the very least, the meaningfulness of the data collected using the BSRI is more limited than the picture frequently painted by researchers' implications.

Bem (1981a) claimed that "[t]he BSRI is...based on a theory about both the cognitive processing and the motivational dynamics of sex-typed and androgynous individuals" (p. 10). She developed these concepts in what became her gender schema theory (Bem, 1981b), proposing that "sex-typing is derived, in part, from a readiness on the part of the individual to encode and to organize information - including information about the self - in terms of the cultural definitions of maleness and femaleness that constitute the society's gender schema" (p. 369). Whether the BSRI is selected as a measure of the degree to which someone is gender-schematic or as a measure of that individual's overall masculinity and/or femininity, however, there are problems inherent in its use. While the BSRI has undoubtedly been useful as an impetus for research and discussion related to gender-role identity and similar constructs (e.g., gender identity, gender role salience, etc.), and its developer has been unequaled in stimulating thought regarding sex-role socialization, it

can be argued that neither a sex-role theory nor an androgyny theory approach to understanding human behavior is any longer efficacious. Bem suggested that individuals can exhibit both "masculine" and "feminine" socially desirable traits. I argue that it is no longer useful to conceptualize traits as such, and I will attempt to demonstrate that, for the most part, the "masculine" and "feminine" items that comprise the BSRI are no longer viewed in gender-linked terms.

Masculinity and Femininity

The lack of meaningfuless in much of the literature on gender, of which the BSRI is only a part, is compounded by the failure of many researchers to distinguish between personality *traits* that have traditionally been defined as "masculine" or "feminine," the *behaviors* that often (but not always) accompany these traits, and the overall *classification of an individual* as highly masculine, feminine, or both (androgynous). For example, a woman may be nurturant (trait) whether she stays home with a child or does not stay home with a child (behavior). Furthermore, her staying home with a child does not mean that she is "feminine," even according to BSRI classifications. Conceptual confusion regarding the phenomenon under investigation, as well as the measures used, has resulted in further contradictory and misleading findings among researchers (Gilbert, 1985).

Individual females (or males) may derive a sense of femininity (or masculinity) from very different things. The elusiveness of the concepts of

masculinity and femininity has been demonstrated in the literature by the difficulty that even highly educated, articulate individuals have in specifying what constitutes their masculinity and femininity (Spence & Sawin, 1985). The basic assumption that underlies most of the research on gender is faulty: that "masculinity" and "femininity" consist of, and are conceptually defined by, lists of traits and interests that are based on sex difference statistics (Ashmore, 1990; Lewin, 1984b).

Masculinity and femininity might best be conceptualized as rather nebulous albeit important aspects of the self-concept that seem to defy definition. If we wish to assess masculinity and femininity without restricting the "richness and diversity" of their meanings for individuals (Marsh & Myers, 1986, p. 428), it may be more fruitful to investigate the individual's *sense of herself or himself* as feminine or masculine.

A case in point may serve to illustrate: an androgynous (BSRI classification) woman may rate herself "7" (always or almost always true) on the BSRI item called "feminine" and "1" (never or almost never true) on the BSRI item called "masculine." Likewise, an androgynous man may rate himself "1" on "feminine" and "7" on "masculine." Many factor analytic investigations of the BSRI have been conducted (reviewed in Lippa, 1985), generally resulting in the conclusion that the scales (Masculine and Feminine) are not factorially pure. Factor analyses of the BSRI typically have yielded two highly correlated instrumentality factors, one of which can be labelled "dominance" and one "self-

reliance"; an expressiveness factor; and a fourth factor often correlated with biological sex, defined by only three of the 60 BSRI items: "feminine," "masculine," and "athletic" (Lippa, 1985). The scenario described above is most likely quite common considering that two ("masculine" and "feminine") or, at best, three items ("masculine," "feminine," and "athletic") have been found to comprise a separate bipolar factor that is orthogonal to the three other factors (Lippa, 1985; Pedhazur & Tetenbaum, 1979). (The inclusion of "athletic" is not surprising when one considers that the development of the BSRI occurred before the passage of Title IX attempted to make athletics non-discriminatory toward females.) It is obvious that "masculine" and "feminine" connote something quite different for BSRI respondents than do items related to instrumentality and expressiveness; what that is, however, remains unknown.

Gender Identity

Gender identity has been described as "a basic, existential conviction that one is male or female" (Spence & Sawin, 1985, p. 59); a secure sense of one's own maleness or femaleness (cf. Green, 1974); and the "individual's awareness of and satisfaction with being a male or female" (Pleck, 1984, p. 220). It would seem that gender identity, defined in these terms, has much more to do with masculinity and femininity than does whatever is being tapped by the BSRI. Lewin (1984b) argued that "[t]here is no evidence that the MF tests of the last sixty years provide a valid measure of the relative femininity of women or the relative masculinity of men" (p. 198). But if masculinity and

femininity are something other than sets of traits and interests, then what are they?

Spence (1985) proposed that "masculinity and femininity, as they refer to an individual's self-concept, be retained and reconceptualized as gender identity: a basic phenomenological sense of one's maleness or femaleness that parallels awareness and acceptance of one's biological sex and is established early in life" (p. 91). Similarly, Lewin (1984b) proposed that masculinity and femininity be conceptualized as "the gender-relevant aspects of a person's self-concept," thus allowing for "individual variation in the specific content of the self-image as related to gender" (p. 200). She suggested that masculinity and femininity measures "should assess gender self-confidence" (p. 200). Indeed, confidence in self has been identified as an important component of global self-concept (Hattie, 1992). Thus, gender self-confidence would seem an appropriate construct for further investigation.

New Theory: Toward a Redefinition of Masculinity and Femininity

Conventional definitions of masculinity and femininity are problematic (Ashmore, 1990; Deaux, 1987; McCreary, 1990; Spence, 1985). Moreover, the widespread interpretation of instruments such as the BSRI as measures of masculinity and femininity leads to muddled and erroneous conclusions. A reevaluation of current theories of gender-role identification and an audit of our "inventory of artifactual and conventional beliefs" (Morawski, 1985, p. 218) is called for.

There are personal costs to men and women when adherence to norms for masculinity and femininity is reinforced (Block, 1973), albeit unintentionally so, through the use of so-called masculinity and femininity measures. These costs take the form of limitations imposed upon both sexes with regard to socially sanctioned behavior. Furthermore, "to assume that scales labeled *masculine* and *feminine* are reliable and valid measures of sex roles, sex-role identity, sex-role orientation, or sex-role beliefs and behaviors is... untenable" (Gilbert, 1985, p. 165).

Masculinity and femininity are much larger constructs than traditional or stereotypical masculine and feminine roles. Researchers who either intentionally or inadvertently reduce the concept of gender identity to the study of gender roles trivialize the importance, and the pervasiveness, of gender identity in human experience. A two-part process is required that involves, first, a closer look at the gender identity concept as a representation of masculinity and femininity, and, second, preliminary steps to assess gender identity as such.

If we continue to ask the wrong questions, we will continue to get wrong answers. Research that focuses on sex differences assumes that there is such a thing as Woman [or Man], and that "Womanness" can be defined in terms of certain qualities (Hare-Mustin & Maracek, 1994). When the ways in which the sexes differ are the focus, differences among women and differences among men are overlooked. As early as 1973, Constantinople argued that the

theoretical explanation that ties sex differences, whatever their content, to masculinity and femininity is lacking. Her landmark work contained the observation that "[i]n all probability, the length of the big toe would discriminate men and women, but does having a longer big toe than most women make a woman less 'feminine'...?" (p. 405.) Androgyny was proposed as a means to ameliorate some of the problems historically associated with the study of masculinity and femininity (Cook, 1987). However, androgyny theory's continued reliance on traditional notions of femininity and masculinity served to reify the very distinction that it sought to blur (Lott, 1981).

Several researchers (e.g., Antill, Cunningham, Russell, & Thompson, 1981; Lewin, 1984b; McCreary, 1990) have commented on the irony inherent in Bem's conceptualization of androgyny as dichotomous masculinity and femininity, and have suggested that it is better not to categorize in the first place. Lewin (1984b) argued eloquently that "nothing productive is accomplished when psychologists first classify traits as either masculine or feminine and then are forced to add hastily 'but of course men are also feminine and women are also masculine' " (p. 197). Bem (1979) herself contended that "behavior should have no gender," and recognized that "the concept of androgyny contains an inner contradiction and hence the seeds of its own destruction.... To the extent that the androgynous message is absorbed by the culture, the concepts of masculinity and femininity will cease to have such content and the distinctions to which they refer will blur into invisibility" (p.

1053).

I suggest that the time has come.

Purpose of the Study

I hypothesize that few meaningful distinctions between masculine and feminine BSRI items currently exist. If this is so, then androgyny can be shown to be the outmoded concept that Bem predicted it would become.

There are some research findings to support this contention. Using a 75% agreement level, Ballard-Reisch and Elton (1992) assessed a predominantly middle-class, Caucasian, *non-college* population in a city located in the western United States for their interpretations of whether the 60 BSRI items were masculine, feminine, or neutral. They found that 19 of the 60 items were viewed as neutral, *only one was viewed as feminine ("feminine"), and only one was viewed as masculine ("masculine")*. Agreement among participants regarding the remaining 39 items was not obtained. A need exists to examine whether similar perceptions are evident among other populations (e.g., an ethnically diverse college population located in the southern United States). If they are, this would suggest that traditional, stereotypical "feminine" and "masculine" characteristics are being integrated and assessed as unlinked to femininity and masculinity, and that the androgyny concept may indeed be obsolete.

Furthermore, there is a need to examine what masculinity and femininity are, as well as what they are not. The literature suggests that masculinity and

femininity can be conceptualized in terms of gender identity. Gender self-confidence has been identified as a construct that merits investigation as a component of gender identity.

Thus, the purpose of the proposed study was to re-examine the androgyny construct and its measurement. The study explored the currency of the BSRI as representative of perceptions held by college undergraduates of "masculine" and "feminine" characteristics. It also examined the concept of gender identity as a representation of masculinity and femininity. Preliminary steps to assess gender identity were taken by examining gender self-confidence as one aspect of gender identity, and devising an instrument designed to assess gender self-confidence. Relationships between participants' levels of gender self-confidence and information obtained using the BSRI were explored.

Need for the Study

Block (1973) and Bem (1993) spoke about the personal costs to men and women when our perceptions and expectations are structured by male and female categories. Such costs stem from an overshadowing of individual human strengths by the amorphous clouds called masculine and feminine. The more that we view the world through the "lenses of gender," and the more we encourage that perspective in our youth, then the more we negate the potential of the human being. We have made it a priority in some circles to downplay differences between racial groups; one wonders why we can't adopt the same

attitude in regard to biological sex. Kindergarten teachers still ask the boys to line up in one area and the girls in another; it would be unacceptable to force that distinction between races.

Counselors need to be aware of gender issues and sensitive to the ways that norms for "masculine" and "feminine" attitudes, traits, and behaviors may have negatively impacted their clients, male or female. As Good, Gilbert, and Scher (1990) argued, counselors must help their clients view personal issues within a societal context, which necessitates consideration of gender dynamics and the potential for harm when gender-related restrictions are knowingly, or unknowingly, imposed.

Counselor educators have many responsibilities to students, counselors to clients, and counseling supervisors to counselors-in-training. One responsibility that seems to be overlooked frequently is the responsibility to critically examine their own gender attitudes and to identify and work to revise biases, stereotypes, and behaviors that might be oppressive to those in their charge. How else will their students, clients, and supervisees learn these same self-assessment skills? Unfortunately, this task is no small one. Well-intentioned helping professionals are not immune to the cumulative effects of years of subtle (and not so subtle) gender-role socialization.

This study was intended as a necessary step toward that end. Meaningful gender research required that we revisit the place we left in the mid-1980s, the need identified by scholars such as Spence (1984, 1985) and Lewin

(1984b): to disentangle and clarify the constructs of masculinity and femininity before we can intelligently discuss them.

Statement of the Problem

The questions addressed by this study had three goals: first, to examine the current viability of the BSRI as a research tool by assessing whether its "masculine" and "feminine" items represent current perceptions of masculinity and femininity among college undergraduates; second, to begin to examine the constructs of masculinity and femininity as representations of an innate sense of gender identity as opposed to umbrellas for stereotypically viewed personality traits; and, third, to assess gender self-confidence as a component of gender identity.

Specifically, this study attempted to answer the following research questions:

1. How do college undergraduates currently describe themselves using the BSRI?

2. Do the "masculine" and "feminine" items on the BSRI represent current perceptions of masculinity and femininity among college undergraduates?

3. What are the major dimensions of college undergraduates' self-reported levels of gender self-confidence?

4. What is the relationship between college undergraduates' levels of gender self-confidence and their self-descriptions according to BSRI

classifications (i.e., masculine, feminine, androgynous, undifferentiated)?

5. What is the relationship between college undergraduates' levels of gender self-confidence and their evaluations of BSRI items as masculine and feminine?

Definitions of Terms

Masculinity and femininity have been defined in a number of ways, as the above indicates. In this study, most of these definitions are being challenged. For the purposes of this study, these two terms refer to the aspects of an individual's self-concept that are gender-relevant to that person (Lewin, 1984b). In this sense, masculinity and femininity can be conceptualized as gender identity (Spence, 1985).

Gender identity refers to a basic phenomenological sense of one's maleness or femaleness that parallels awareness and acceptance of one's biological sex (Spence, 1985).

Gender self-confidence refers to that aspect of gender identity related to one's self-assuredness about being male or female.

Sex is a biological term that defines people as male or female depending on their organs and genes (Mintz & O'Neil, 1990).

Gender roles refer to behaviors, expectations, and roles defined by society as stereotypically linked to males (masculine) or females (feminine); also referred to as societal gender roles (Mintz & O'Neil, 1990).

Gender role socialization refers to the process by which people in a

particular culture are taught about societal gender roles (Mintz & O'Neil, 1990).

Gender role identity, also referred to in the literature as sex-role identity or sex-role orientation, refers to the degree to which a person identifies with or displays societally defined masculine or feminine behavior (Basow, 1992; Mintz & O'Neil, 1990).

Sex-typed refers to an individual who is classified as "masculine" or "feminine" according to the Bem Sex-Role Inventory (BSRI) in accordance with his or her biological sex. In order to be classified as "masculine" on the BSRI, a male must score higher than the median on the "masculine" dimension and lower than the median on the "feminine" dimension. In order to be classified as "feminine" on the BSRI, a female must score higher than the median on the "feminine" dimension and lower than the median on the "masculine" dimension.

Cross-sex-typed refers to a male who is classified as "feminine" according to the BSRI (scoring higher than the median on the "feminine" dimension and lower than the median on the "masculine" dimension) or a female who is classified as "masculine" according to the BSRI (scoring higher than the median on the "masculine" dimension and lower than the median on the "feminine" dimension).

Androgynous refers to an individual who scores higher than the median on both the "masculine" and "feminine" dimensions of the BSRI.

Undifferentiated refers to an individual who scores lower than the median on both the "masculine" and the "feminine" dimensions of the BSRI.

Overview of Remaining Chapters

Chapter II provides a more detailed analysis of the history and current status of the topic of this study. It includes discussion of the history of masculinity and femininity measurement, the revolutionary work of Sandra Bem and other researchers of the "androgyny era," the use of the BSRI in research, critiques of the BSRI, and an overview of the gender identity construct. Relevant theoretical and empirical research is reviewed. The methodology used in the study is described in Chapter III, including procedures for development of the instrument. Chapter IV provides a comprehensive account of the results of data analyses. Conclusions, implications, and recommendations for further study are discussed in Chapter V.

CHAPTER II

REVIEW OF RELATED LITERATURE

The literature pertinent to this study involves five areas: (a) the history of masculinity and femininity measurement through 1970, (b) the revolutionary work of Sandra Bem and others, (c) use of the Bem Sex-Role Inventory in research, (d) critiques of the BSRI, and (e) the gender identity construct. Following a review of all five areas, a rationale for the proposed study based on the review is presented.

Measurement of Masculinity and Femininity (1936-1970)

Like most psychological constructs, masculinity and femininity are abstract concepts. Yet the lack of success in measuring masculinity and femininity over the last 60 years makes these constructs more elusive than most (Constantinople, 1973; Lewin, 1984b; Spence, 1993).

Until the mid-1970s, it was commonly believed that masculinity-femininity was bipolar and unifactorial (Bem, 1981a; Spence, 1993). By this it is meant that masculinity and femininity were conceptualized as opposite ends of a single continuum along which every individual could be placed. Masculinity-femininity was thought to be best defined in terms of sex differences in item responses. Furthermore, all of the psychological traits that were said to distinguish between men and women were seen as part of one aggregate

labelled Masculinity-Femininity (Spence, 1993). From the turn of the century until the 1970s, psychologists struggled to quantify "MF" using these assumptions.

Terman and Miles: Attitude-Interest Analysis Survey (AIAS)

From the mid-1930s through the mid-1950s, the study of sex and gender entailed the introduction and acceptance within the psychological arena of the notion of a pair of general and opposing personality traits identified as "masculinity" and "femininity" (Ashmore, 1990). The most notable work of the time was a book entitled Sex and Personality by Terman and Miles (1936). Intelligence testing provided a model for their development of a self-report measure of "masculinity-femininity." This instrument was named the Attitude Interest Analysis Survey (AIAS; Terman & Miles, 1936) in order to reduce the possibility that responses might be influenced by a knowledge of the scale's purpose. As such, incongruities between one's biological sex and one's "psychological" sex were thought to be identified (Morawski, 1987). The AIAS provided a bridge to the measurement of homosexuality, described by psychologists of the time as "sexual inversion," based on the conviction that feminine women and homosexual men must have a lot in common (Lewin, 1984a; Morawski, 1987). The AIAS was further believed to be useful in predicting problems in marital adjustment by identifying "feminine" and "masculine" minds (Morawski, 1987). According to Terman and Miles, "mental masculinity and femininity" was at the core of an individual's temperament and

provided the basis for the rest of the personality.

Constantinople (1973) noted that Terman and Miles "offered no definitions of the trait which are grounded in theory" (p. 392). Masculinity and femininity were defined by Terman and Miles purely in terms of sex differences in response (Constantinople, 1973). Although the instrument was intended for use with adults, it was normed on students in elementary through high school, with items assigned a classification as "feminine" or "masculine" based on average differences in response between girls and boys (Lewin, 1984a). Terman and Miles themselves acknowledged deficiencies in both the adequacy of the criterion and the measurement process they used in the construction of the 456-item AIAS (Constantinople, 1973).

Of the seven subtests, Terman and Miles considered "Emotional and Ethical Attitudes" and "Interests" to be the strongest and most reliable (Lewin, 1984a). Racial as well as gender bias is evident in certain test items (e.g., "Negroes" was selected as a word intended to arouse fear in respondents). The fact that scoring of the instrument involved awarding "plus" for "masculine" responses and "minus" for "feminine" responses carries its own subtle message.

Despite their own misgivings about their attempt to measure "M-F," the work of Terman and Miles (1936) became the pattern for subsequent research in masculinity and femininity measurement.

Masculinity-Femininity Scale of the Strong Vocational Interest Blank (SVIB)

The purpose of the Strong Vocational Interest Blank (SVIB; Strong, 1927) was to identify occupations in which an individual might share similar interests with others choosing the same occupations. The SVIB MF scale was intended to discriminate between typically masculine and typically feminine occupationally-related interests. Like the other scales of the SVIB, the MF scale was constructed based on weighting responses proportionately to the weight assigned by criterion groups (Lewin, 1984a). While the Terman-Miles M-F measure included only those test items on which there were significant differences between women and men, the SVIB MF scale included all items that showed any differentiation. Contrary to his original position, however, by 1943 Strong acknowledged that, in general, similarities in interests between the sexes were much stronger than differences (Constantinople, 1973; Lewin, 1984a). Because males and females agreed on 86.5% of the items, the MF scale was comprised of only 13.5% of the items Strong used (Lewin, 1984a). Like Terman, his predecessor and mentor, Strong identified differences between the sexes as the criterion for measurement of masculinity and femininity, and accepted the assumption of bipolarity.

Minnesota Multiphasic Personality Inventory (MMPI) Masculinity-Femininity Scale

The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) was originally developed as an assessment tool for use with

individuals suspected of exhibiting some degree of psychopathology (Thorndike & Hagen, 1977). The Masculinity-Femininity scale (Mf) focused on "persons tending to identify with the opposite sex, rather than their own" (Thorndike & Hagen, 1977, p. 425). High scores on the Mf scale indicated "femininity," which has been interpreted as "probably sensitive and idealistic with high aesthetic, cultural, and artistic interests" (Thorndike & Hagen, 1977, p. 426).

Those who are familiar with the test development procedures of the MMPI Mf scale agree that scores derived from its use should be viewed with some concern (Constantinople, 1973; Lewin, 1984b, 1991). Unfortunately, most psychologists and counselors, let alone "untrained" people who have access to individuals' MMPI profiles (e.g., human resource managers, etc.), are not aware that the femininity dimension was originally "validated" on a criterion group of 13 male homosexuals (Lewin, 1984b). Unfortunate also is the fact that as one of the 10 basic clinical scales of the MMPI, the Mf scale is one of the most widely used in research, counseling, and job screening.

It should be noted that the Mf scale of the MMPI-2, released in 1990, differs from the original in that four of the 60 items were dropped as "potentially offensive" (Lewin, 1991, p. 585). The curious reader would, of course, consult the test manual to find which four items this statement is describing; however, the test manual does not divulge this information. Rather, it states what Lewin and Wild (1991) had reiterated, that four of the original Mf items were deleted because of "objectionable content" (Butcher, Dahlstrom, Graham, Tellegen, &

Kaemmer, 1989, p. 29). A laborious comparison of the original MMPI and the MMPI-2 revealed that one of the four items that was omitted from the revised MMPI Mf scale was "I used to like drop the handkerchief." Why such blatantly stereotypical items as "I think I would like the work of a librarian" or "I like mechanics magazines" were not thought to comprise "objectionable content" when a merely antiquated one is deemed such is not clear.

Current scale score distributions, means, and T-scores were obtained using a new normative sample intended to be representative of the United States population. The MMPI-2 (Butcher et al., 1989) contains two new scales identified as the Masculine Gender role scale (Gm) and the Feminine Gender role scale (Gf), intended to tap traditionally masculine and feminine gender roles. Items included in these scales were selected from items on the old Mf scale (Butcher et al., 1989). The Gm and Gf scales are not among the 10 basic clinical scales, so they are not always scored. Progress toward equality appears absent in light of Lewin and Wild's (1991) observation that males get masculinity points for responding "true" to the following items: "I am worried about sex," "I like to talk about sex," and "I wish I were not bothered by thoughts of sex," whereas females receive femininity points for responding "false" to these statements (p. 586).

The test manuals contain no clear definition of what the Mf scale of the original MMPI or the MMPI-2 is designed to measure. What is clear is that homosexuality in males is confused with femininity on more than one occasion.

In addition to gay males being used as a validating criterion on the original MMPI, users of the MMPI-2 are advised by the manual that males who score highly "feminine" are likely to be "passive," "empathic," and "have homoerotic trends" (Butcher et al., 1989, p. 38). Further perusal of the MMPI-2 manual (Butcher et al., 1989) reveals that males who score very high on the Mf scale also may have "conflicts over sexual identity" (p. 38). It seems safe to conclude that neither the MMPI Mf scale nor the MMPI-2 Mf scale is an adequately validated measure of masculinity and/or femininity (Constantinople, 1973; Cronbach, 1960; Lewin and Wild, 1991).

The GAMIN Inventory Masculinity Scale

Between 1936 and 1956, J. P. Guilford and his associates applied factor analytic procedures to determine dimensions of introversion-extraversion as part of a continuing attempt to discover basic dimensions of personality. Although they questioned whether one of the factors represented a masculine ideal or dominance rather than a sex-difference factor, the decision was made to identify it as a masculinity factor and to include it in the GAMIN inventory (General activity, Ascendance vs. submission, Masculinity vs. femininity, confidence vs. Inferiority feelings, and calmness vs. Nervous) (Guilford & Martin, 1943). Factor M was incorporated into the Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1949). Caution regarding labeling the factor "M" had dissipated by 1956, and although mention was made of the possibility that MF and sex differences might not be the same thing, the

subject appeared to have been dropped (Lewin, 1984a). As a result, "M" came to be measured by 40 items comprising six subtests, named Inhibition of Emotional Expression, Masculine Vocational Interests, Masculine Avocational Interests, Disgustfulness, Fearfulness, and Sympathy (Lewin, 1984a). The classic "cult of true womanhood" (Welter, 1978) was evident in the latter three scales, pseudonyms for Purity, Submissiveness, and Moral Superiority, respectively (Lewin, 1984a).

Gough: The Femininity Scale (Fe) of the California Psychological Inventory (CPI)

The Fe scale of the California Psychological Inventory (CPI; Gough, 1952) was intended to differentiate males from females and "sexual deviates from normals" (p. 427). Item clusters were similar to those found in other M-F tests (e.g., sensitivity to social interaction, social timidity and lack of confidence, compassion and sympathy) and were generally representative of gender stereotypes (Constantinople, 1973). Although moderate correlations were found between the CPI Fe scale and the SVIB MF scale (-.41) and the MMPI Mf scale (+.43) (Gough, 1964), construct validity remains lacking due to the fact that a considerable proportion of the variance of any two of the tests described is not held in common (Constantinople, 1973).

General Critiques of MF Measures

Anne Constantinople (1973) was the first major scholar to comprehensively review existing masculinity and femininity measures. She

focused on what she identified as three untested assumptions related to the nature of the M-F construct:

(a) the assumption that masculinity and femininity are best defined in terms of sex differences in responses, (b) the assumption that a single bipolar dimension exists ranging from extreme masculinity at one end to extreme femininity at the other, and (c) the assumption that the masculinity-femininity construct is unidimensional in nature and is appropriately measured by a single score (Constantinople, 1973).

The last assumption concerned the possibility of multidimensionality, or the existence of sets of traits or factors that could be identified, rather than the issue of measuring masculinity and femininity separately, which is in the realm of the bipolarity assumption. Constantinople cited the work of Webster (1956), who identified conventionality (preference for conventionally feminine roles and interests), passivity, and sensitivity as three factors that discriminated the sexes in her study of Vassar College students. Webster's finding that women in college became more "masculine" in the sense of becoming less conventional and less passive, but more "feminine" in the sense of becoming more sensitive and introspective would have been lost if she had looked at the total score only and ignored the subscores. Constantinople (1973) gave numerous other examples of how studies using contemporary MF tests, namely the ones described above, showed evidence of a variety of factors as well.

The second assumption, that of bipolarity, apparently had widespread

appeal to those who thought dichotomously. However, as early as 1953, Webster contended that "psychological femininity" may be considered as a personality variable that is present in both sexes. Although she did not address "masculinity" per se, her arguments were consistent with the premises of Jungian theory, suggesting that this was an area that merited further thought (Constantinople, 1973). Furthermore, bipolarity implies that there should be close to a (-1.0) correlation between M and F, when, in fact, data from studies using contemporary M-F instruments indicated more positive than negative correlations between items selected to represent masculinity and those selected to represent femininity.

The first assumption, claiming that establishing differences between males and females in their responses to test items is a viable criterion for determining measures of masculinity and femininity, may pose the biggest challenge to researchers. As Anne Contantinople (1973) so eloquently stated, "The universe of sex differences is large indeed, and it is not unreasonable to expect that these differences reflect more than one underlying dimension" (p. 398). Her classic illustration of the absurdity of basing MF measures on sex differences follows:

> While it is clear that something is being measured by the tests of M-F...the theoretical explication that would tie sex differences...to masculinity and femininity is absent.... [T]he length of the big toe would discriminate men and women, but does having a longer big toe than most women make a woman less "feminine," and can one have more confidence that she is less "feminine" because she scores deviantly on a number of items with similarly critical content? (p. 405)

But if sex differences are not an adequate criterion, then what should the criterion be? Constantinople (1973) suggested that this assumption is certainly open to question, but speculated that it would be the most difficult to address.

By and large, researchers who responded to Constantinople's criticisms of MF studies focused on the challenge to bipolarity and ignored the other two assumptions (Marsh & Myers, 1986). Procedures used to develop MF instruments have been largely atheoretical, offering a weak basis for the development or refinement of any theory (Marsh & Myers, 1986). Difficulty notwithstanding, the critical issues of multidimensionality and establishing a meaningful criterion for MF evaluation cannot be ignored. Without a meaningful criterion, it would seem that additional research is superfluous.

Similarly, Lewin (1984a) presented a list of what she identified as eight assumptions unwittingly but mistakenly made by MF test developers prior to 1970. These eight assumptions (slightly reworded below) merit enumeration:

1. None of the tests was validated as a measure to differentiate between more and less feminine women, or more and less masculine men, although that is identified as an objective of an MF test. Within-sex validity studies were non-existent.
2. Any "appealing" items that showed sex differences were accepted as measures of femininity or masculinity, no matter how irrelevant they might be.
3. Femininity and masculinity were assumed to be opposite ends of a

single continuum or dimension.

4. MF was conceptualized as a static trait unaffected by developmental processes; therefore, MF tests were constructed using children as criterion groups.
5. The responses of gay men and feminine women were treated as identical.
6. Advocates of certain projective MF tests [e.g., the Franck Test (Franck & Rosen, 1949)] assumed that substantial proportions (28% - 40%) of nonpatient general populations had an unconscious other-sex gender identity (i.e., were transsexual) based upon their interpretations of subjects' drawings.
7. Ignored was the fact that sex-role or gender (MF) norms are intimately linked to the economic, political, and social conditions in society and therefore change with those conditions. It was tacitly assumed that the MF ideals of 19th century Victorian America were universal and fixed.
8. MF was conceptualized as a set of human traits and interests. No allowance was made for individual variation in the gender-related content of the self-image. Therefore, MF as an aspect of the self-concept was ignored (pp. 167-168).

Lewin (1984a) argued that of all these faulty assumptions, the final one might merit the greatest attention as perhaps the most fundamental. Lewin

(1984b) proposed that MF be conceptualized as "the gender-relevant aspects of a person's self-concept or self-image," allowing "room for individual variation in the specific content of the self-image as related to gender" (p. 200).

The sets of questionable assumptions underlying measures of masculinity and femininity pointed out by Constantinople (1973) and Lewin (1984a) share several commonalities. Most notably, these include the assumption of bipolarity, as well as the assumption that using sex differences in responses is an acceptable criterion for measuring masculinity and femininity. Although Constantinople (1973) expressed more criticisms related to the assumption of the unidimensionality of both masculinity and femininity than did Lewin (1984a, 1984b), Lewin focused in greater detail on problems stemming from conceptualizing femininity and masculinity in a restrictive way, as well as other issues pertaining to construct validity.

Constantinople's (1973) work has been praised by researchers regardless of their theoretical perspectives or backgrounds. It has been cited by Ashmore (1990) as a "detailed and well-reasoned critique" (p. 503), described by Marsh and Myers (1986) as a "classic review of Masculinity-Femininity [MF] research" with "convincing evidence that MF is multidimensional" (p. 398), and characterized by "persuasive evidence that these tests had artificially constrained the relationship between masculinity and femininity" (Lenney, 1991). Similarly, Lewin's work (1984a, 1984b; Lewin & Wild, 1991) has been lauded by eminent scholars in the field (cf. Ashmore,

1990; Morawski, 1987).

Constantinople (1973) posed another thought-provoking question in her landmark work: "If M-F reflects a number of subtraits, such as aggressiveness, sensitivity, self-confidence, etc., is there anything to be gained by combining these measures in ways that are most characteristic of men and women?" (p. 405).

Second Stage Theories: The Bem Sex-Role Inventory and Other Revolutionary Work in the Field of Masculinity and Femininity Research

The 1970s heralded a new concept in masculinity and femininity research: the idea that healthy males and females could possess similar characteristics. Androgyny emerged as a framework for interpreting similarities and differences among individuals according to the degree to which they described themselves in terms of characteristics traditionally associated with men (masculine) and those associated with women (feminine) (Cook, 1987). Although the term "androgyny" is not new, having its roots in classical mythology and literature (andro = male, gyne = female), the 1970s marked a resurgence of the word's popularity as a means to represent the combination of personality traits stereotypically associated with one or the other sex.

The Bem Sex-Role Inventory (BSRI; Bem, 1974) was designed to facilitate empirical research on psychological androgyny. The BSRI differed from earlier instruments in that its developer challenged the assumption of bipolarity and theorized that the constructs of masculinity and femininity are

conceptually and empirically distinct. The construction of the BSRI included a separate Masculine scale and a separate Feminine scale, which Bem defined in terms of culturally desirable traits for males and females, respectively. She argued that an individual could possess a number of traits from each scale, and that one could demonstrate varying degrees of such traits in response to different situations.

The BSRI consists of 60 personality characteristics on which respondents are asked to rate themselves on a 7-point Likert scale, ranging from 1 (Never or almost never true) to 7 (Always or almost always true). Twenty of the characteristics are stereotypically feminine (e.g., affectionate, sympathetic, gentle), 20 are stereotypically masculine (e.g., independent, forceful, dominant), and 20 are considered filler items by virtue of their gender neutrality (e.g., conscientious, conceited, truthful). These 20 neutral items were used to comprise a measure of Social Desirability in response. Unlike the "feminine" items and the "masculine" items, which were all identified as socially desirable for their respective sex, 10 of the gender-neutral items were identified as desirable for both sexes (e.g., adaptable, sincere) and the other 10 as undesirable for both sexes (e.g., inefficient, jealous).

When the BSRI was first published, scoring procedures and interpretation were such that if an individual's Femininity raw score exceeded his or her Masculinity raw score at a statistically significant level, the respondent would be classified as "feminine"; if the reverse were true, the

individual would be labelled "masculine"; and if the difference were small and not statistically significant, that person would be called "androgynous." Spence, Helmreich, and Stapp (1975) pointed out that this process did not differentiate between those who scored low on both scales and those who scored high on both scales. To correct this deficiency, Bem (1977) proposed a modification of scoring that resulted in the current procedure of a median-split to form four distinct groups: feminine, masculine, androgynous, and undifferentiated. A difference score (between femininity and masculinity) is determined based on standardized T-scores. Bem suggested that researchers might use the median-split classification derived from their own research populations if their samples are large and comprised of both males and females, and that they may wish to utilize the medians of the normative sample if working with a small or single-sex sample.

Either way, the median-split classification system allows the respondent to ascertain whether he or she rates *high* on both dimensions (masculinity and femininity), thus classified as "androgynous," *low* on both dimensions ("undifferentiated"), or high on one dimension but low on the other (sex-typed as either "masculine" or "feminine" if the high-scoring dimension corresponds to the person's sex, or cross-sex-typed if the low-scoring dimension corresponds to one's sex).

Soon after the development of the original version of the BSRI, Bem (1979, 1981a) constructed the BSRI Short Form. It contains 30 of the original

60 items, with 10 items comprising each of the three scales (Masculinity, Femininity, Social Desirability). Bem's purpose in developing the Short Form of the BSRI was to address concerns related to poor item-total correlations with the Masculinity and Femininity scales as well as issues raised by factor analyses (Lenney, 1991). These issues are discussed in a subsequent section of this chapter.

Bem (1981a) contended that the BSRI is "based on a theory about both the cognitive processing and the motivational dynamics of sex-typed and androgynous individuals" (p. 10). These concepts, briefly referred to in the test manual (Bem, 1981a) provided the basis for the development of Bem's gender schema theory (Bem, 1981b, 1981c). The main tenet of gender schema theory is that "sex-typing is derived, in part, from a readiness on the part of the individual to encode and to organize information - including information about the self - in terms of the cultural definitions of maleness and femaleness that constitute the society's gender schema" (p. 369). According to Bem (1987), a sex-typed individual is someone whose self-concept incorporates prevailing cultural definitions of masculinity and femininity.

Bem's instrument was the first test specifically designed to provide independent measures of an individual's masculinity and femininity (Lenney, 1991). Bem's distinct purpose was "to assess the extent to which the culture's definitions of desirable female and male attributes are reflected in an individual's self-description" (Bem, 1979, p. 1048). Thus, she defined

masculinity and femininity in terms of sex-linked social desirability.

The BSRI and gender schema theory spurred cataclysmic changes in the way femininity and masculinity were conceptualized. For the first time, masculinity and femininity were defined "from the outside in" (Ashmore, 1990). This meant that the ramifications of gender at the societal level were acknowledged as critical to the formulation of individual self-definitions. This perspective was very different from that of Terman and Miles (1936) and other M-F test developers of the previous era, who had ignored the cultural context and focused exclusively on differences in responses between the sexes to determine what was "feminine" and what was "masculine." Furthermore, Bem's work redefined the relationship between psychological health and gender. The assumption that it was healthy for individuals to be sex-typed was replaced by the assertion that traditionally "feminine" and traditionally "masculine" qualities could be healthy regardless of one's biological sex. While sex differences had come to be minimized, however, the words "masculine" and "feminine" were now maximized as labels for specific characteristics.

At the same time that Bem was developing the BSRI, another pioneer in gender research, Janet Spence, was working with two of her colleagues on another revolutionary instrument, the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974). Spence had focused primarily on the assessment of gender stereotypes and related concepts, such as gender attitudes, as evidenced by the development of the Attitudes Toward Women

Scale (AWS; Spence & Helmreich, 1972). In contrast to the BSRI, whose Masculine and Feminine scales included only items judged to be significantly more desirable for one sex than the other, the PAQ included items judged to be desirable for *both* sexes but seen as more *typical* of one sex than the other. The procedure used by Spence et al. (1974), therefore, focused on the desirability attached to certain traits independent of gender, while still acknowledging the difference between social ideals and reality. The PAQ further differed from the BSRI in that instead of consisting of only two scales, a third scale (M-F) was developed to represent those characteristics whose social desirability varied according to whether an individual was male or female. Although Spence and Helmreich (1978) acknowledged some embarrassment at simultaneously embracing a dualistic and a bipolar model of masculinity and femininity, they retained the M-F scale as a source of "significant information not available from the other scales" (p. 20).

Perhaps the most noteworthy point of departure between the BSRI and the PAQ is what each is purported to measure. Spence has repeatedly argued that the BSRI *and* the PAQ are basically measures of instrumentality and expressiveness (Spence, 1993; Spence & Helmreich, 1981). Bem (1981b), on the other hand, has suggested that the BSRI does indeed tap masculinity and femininity as constructs via the assessment of gender-schematic processing. This issue will be discussed in greater detail in the section of this chapter that deals with critiques of the BSRI.

A lesser known but equally valuable perspective was presented by Hefner, Rebecca, and Oleshansky (1975) in their work on sex-role transcendence. Their model is described as a progression through three stages: (a) an early childhood unawareness of culturally imposed gender-linked restrictions on behavior, (b) a polarized, oppositional view of sex roles which develops in childhood and is largely maintained in adulthood, and (c) a dynamic transcendence of conventional sex roles that involves a "reorganization of the possibilities learned in Stage II in a more personally and socially relevant framework" (Hefner et al., 1975, p. 151). The third stage is conceptualized as the beginning of a dialectical orientation to life rather than the end of a process. Unfortunately, Stage III represents a state rarely achieved because, unlike the shift from the first to the second stage, there is virtually no societal support for the transition from Stage II to Stage III. According to Meda Rebecca, who further developed the concept of sex-role transcendence (Rebecca, Hefner, & Oleshansky, 1976; Rebecca & Hefner, 1979), individuals in Stage III are free to express their human qualities without retribution for violating sex-role norms. This consequence, or lack thereof, would imply changes that involve more than the individual; sex-role transcendence speaks to changes in the larger society as well. In contrast to Bem's perspective, sex-role transcendence goes beyond situational flexibility; here, one does not feel the need to compromise one's personal integrity by adopting an aggressive style, for example, in order to succeed in a professional role. Rather, the role itself might be altered to

accommodate alternative behaviors. Rebecca et al. (1976) contended that Bem's work went "part way toward the dynamic and flexible conception of transcendent Stage III, but still within a trait model" (p. 205). Garnets and Pleck (1979) and Ravinder (1987) were among those who supported and utilized the sex-role transcendence model.

The challenges to traditional perspectives in MF measurement, particularly those voiced through the work of Bem and Spence, dramatically changed the way researchers approached the study of gender. Masculinity and femininity were finally recognized as two independent dimensions. Furthermore, characteristics generally associated with one sex or the other came to be viewed as healthy for both sexes to possess and to demonstrate. Interest in the concept of androgyny skyrocketed, although its origins in classical mythology and literature verified its status as an ancient rather than a new concept (Cook, 1985). Despite the fact that the concept of androgyny encouraged thinking beyond that which was stereotypically masculine or feminine, it still encouraged individuals to perceive certain traits as masculine or feminine by labelling them as such. This is precisely why the notion of sex-role transcendence (Hefner et al., 1975; Rebecca & Hefner, 1979, Rebecca et al., 1976), which encouraged the omission of masculine and feminine labels, was even more revolutionary and possibly healthier as well. Unfortunately, the concept of sex-role transcendence was perhaps too "evolved" for its time, as it did not receive the same type of attention in the literature that androgyny did.

But androgyny certainly can be viewed as movement in a positive direction, and the work of Bem and Spence spearheaded this movement.

Because Bem's stated purpose in the development of the BSRI was to facilitate empirical research on psychological androgyny, and because her instrument is the most widely utlized by MF researchers, her work will be the primary focus of the discussion that follows.

The Widespread Use of the Bem Sex-Role Inventory in Research

The Bem Sex-Role Inventory (BSRI) is the most commonly used measure in all areas of gender-related research (Beere, 1990). A literature search conducted by Beere (1990) in preparation for her anthology of gender tests and measures identifed 795 articles and 167 ERIC documents that used the BSRI. None of those references was a duplicate of those listed in her first book (Beere, 1979).

The BSRI has been used most extensively with college students (Beere, 1990). Considering that this is the group on which the instrument was primarily normed, this hardly seems inappropriate. According to Beere (1990), the BSRI also has been administered to:

> [professional] athletes, physicians, attorneys, hotel employees, married couples, adolescents, infertile couples, parents-to-be, parents, senior citizens, college faculty, women awaiting trial, counselors-in-training, public school administrators and educators, medical, dental, and dental hygiene students, incarcerated criminals, women with gynecological problems, anorexics and bulimics, middle managers, ministerial students and ministers, nursing students, psychiatric inpatients, meditators, teachers, psychotherapists, high school and college athletes,

> homosexual fathers, patients, physicians, career counseling clients, juvenile delinquents, physical educators, clinical psychologists, university faculty, police cadets, athletic administrators, health professionals, accountants, museum visitors, women receiving abortions, social workers, prostitutes, alcoholics, and schizophrenics. (pp. 74-75)

In addition, both the long and short forms of the BSRI have been used in a variety of countries including the United States, Germany, New Zealand, Australia, Israel, India, West Indies, Ireland, Sweden, South Africa, Canada, Mexico, Saudi Arabia, Malaysia, and Finland (Beere, 1990). The BSRI has been modified by many researchers, some slightly (e.g., Wheeless & Dierks-Stewart, 1981) and others extensively, some resulting in new measures for additional cross-cultural research (e.g., Personal Description Questionnaire; Antill, Cunningham, Russell, & Thompson, 1981). Other more extensive modifications have resulted in assessment tools for use with younger populations. These include the Adolescent Sex Role Inventory (ASRI; Thomas & Robinson, 1981), the Children's Sex Role Self-Concept Inventory (Kurdek & Siesky (1980), and the Children's Sex Role Inventory (CSRI; Boldizar, 1991).

A series of studies was conducted by Bem and her colleagues (e.g., Bem, 1975, 1977, 1979; Bem & Lenney, 1976; Bem, Martyna, & Watson, 1976) to establish validity of the original BSRI. These, as well as some of the major factor analytic studies of the BSRI (e.g., Gaudreau, 1977; Martin & Ramanaiah, 1988; Pedhazur & Tetenbaum, 1979), are discussed in the following section.

Despite its unequivocal popularity as a research tool, the appeal of the BSRI is far from universal. It has been attacked relentlessly, as has its developer, in relation to conceptual as well as methodological issues (e.g., Pedhazur & Tetenbaum, 1979). Ironically, these attacks have contributed to the BSRI becoming even more widely known as an MF measure, and, consequently, even more utilized by researchers. In fact, it would seem that the BSRI has been repeatedly used without clear and deliberate thought to the research questions being studied (Gilbert, 1985), the blame for which is hardly attributable to the test developer. It is to the critiques of the BSRI that we now turn.

Critiques of the Bem Sex-Role Inventory

The plethora of androgyny research using the BSRI yielded many inconsistent findings and failures to replicate (for in-depth reviews see Ashmore, 1990; Cook, 1985). At the outset, however, it must be said that it is difficult to develop a really sound measurement tool; it is far easier to critique one already in existence. With this in mind, the Bem Sex-Role Inventory is examined in relation to its theoretical basis, item selection procedures, score interpretation, construct validity, reliability, and factor analysis/dimensionality.

Theoretical Rationale

Gender schema theory already has been discussed as the foundation upon which the BSRI was constructed, although it was not defined as such until several years later. Bem (1981b) referred to the process by which a society

"transmutes male and female into masculine and feminine" as "sex-typing" (p. 354). She contended that gender schema theory explained why sex-typed individuals process information in gender-linked terms and non-sex-typed persons do not. It became more evident in some of her subsequent work (Bem, 1985, 1987) that her theory speaks as much to society's gender schema as it does to the individual's. The problem with Bem's perspective here is that it contains an implicit assumption that culture is homogeneous, which it is not. Not only are there variations within a culture (e.g., American society), but individuals within that society also do not receive consistent messages from all the components of that system (Ashmore, 1990). Furthermore, Bem seemed to conceptualize an individual as a "passive recipient of societal forces" (Ashmore, 1990, p. 507).

Even if "cultural definition[s] of maleness and femaleness that constitute[d] the society's gender schema" (Bem, 1981c, p. 369) did exist, it has been argued repeatedly that "maleness" and "femaleness" are quite different from the more limited notion of traditional male and female *roles* (Spence, 1985). Furthermore, Bem herself has reconsidered the concept of androgyny and found it to be problematic because of its presupposition that the constructs of masculinity and femininity have "an idependent and palpable reality rather than themselves being cognitive constructs" (Bem, 1985, p. 221). Bem's original premise when she developed the BSRI was turned around in 1985 when she contended that "[i]n short, human behaviors should no longer be

linked with gender, and society should stop projecting gender into situations irrelevant to genitalia" (Bem, 1985, p. 222). More recently, Bem (1993) cautioned readers to resist the lenses of gender that structure our perception of the world in female and male categories, thereby imposing severe limitations upon both sexes. In light of Bem's current thinking, one cannot refrain from questioning the implications of past and current usage of the BSRI.

Spence and Helmreich's (1981) analysis of the BSRI led them to conclude that the instument is basically a measure of instrumentality and expressiveness. The debate continued with Bem's (1981b) response that "the empirical evidence demonstrates that the very act of describing oneself as sex-typed on the BSRI is, in part, a product of gender-schematic processing," which Bem contended

> ...reflects different things for different people. For non-sex-typed individuals, the BSRI may well tap [only] instrumental and expressive traits...describ[ing] themselves as, say, dominant or nurturant without implicating the concepts of masculinity and femininity. When sex-typed individuals do describe themselves, however, it is precisely the masculine/feminine connotations of the items on the BSRI to which they are responding. (p. 370)

In interpreting Bem's full reply (Bem, 1981b), Spence (1991) identified implicit support rather than the intended refute of Spence and Helmreich's contentions that "the PAQ and the BSRI are basically personality inventories, and that neither is an acceptable predictor of sex-role attitudes, behaviors, or preferences; hence, they are deficient as measures of broad gender concepts

such as sex-role identification" (p. 159).

An additional source of confusion is Bem's discussion of gender schema as an "all or nothing" phenomenon. That is, she addresses gender schema theory as accounting for an individual's processing of information in gender-linked terms, defined as masculine *and* feminine. This perspective does not allow for the possibility that an individual might be predisposed to interpret information in either masculine *or* feminine terms, but not the other, a position espoused by Markus, Crane, Bernstein, and Siladi (1982).

Item Selection Procedures

The item selection procedures used by Bem in the construction of the BSRI were designed to assess "not what particular members of a given culture themselves define as masculine or feminine but what they collectively believe to be the prevailing definitions of masculinity and femininity in the culture at large" (Bem, 1987, p. 267). This approach is consistent with gender schema theory, which holds that it is the collective cultural definitions that the sex-typed person uses as the criteria for his or her gender conformity (Bem, 1987).

However, confusion has resulted from Bem's (1981c) indiscriminate use of the terms "masculinity" and "femininity" as constructs measured by the BSRI. On the one hand, Bem allows the individual to have personal definitions of masculinity and femininity, yet holds these definitions as irrelevant to gender-schematic processing and sex-typing (Ashmore, 1990). It does seem odd that a measurement tool comprised of items selected for their sex-specific

desirability can be used to validate a theory which purports that males and females are free to have attributes from both the "masculine" and "feminine" domains (McCreary, 1990). Here again, one wonders why traits must be classified as either masculine or feminine when the caveat that "men are also feminine and women are also masculine" is inevitably attached (Lewin, 1984b, p. 197).

In addition to the conceptual confusion that characterizes BSRI item selection procedures, methodological problems also are evident (Pedhazur & Tetenbaum, 1979). Bem (1974) used independent t-tests to ascertain whether each of the 400 items in her pool of adjectives was significantly more desirable for a man than for a woman (to qualify as masculine), or for a woman than a man (to qualify as feminine). According to Pedhazur and Tetenbaum (1979), this easily could have resulted in the inclusion of items as "masculine" and "feminine" that were judged as not necessarily more desirable for one sex than the other, but rather as *less undesirable* for one of the sexes. The fact that items such as "gullible" qualified as "feminine" using these procedures seems to support this observation. As noted in Chapter I, this would appear to be an example of statistically significant results overshadowing substantively meaningful findings.

The persons who served as judges of the social desirability of potential BSRI items were 100 undergraduate students at Stanford University in 1972. Fifty of the students were females and 50 were males. Judges were asked to

rate the desirability of each adjective *either* "for a woman" *or* "for a man"; no judge rated desirability of these characteristics for both women and men (Bem, 1981a, p. 11). Although the manual seems to suggest that Bem initiated this procedure in an attempt to strengthen test construction, it may be that the result was the opposite. Bem's (1981a) procedures provided no way of comparing how a judge would rate the desirability of an item for a female versus a male.

Attempts have been made to replicate item selection procedures for the BSRI, some with modifications and others without. In general, the purpose of such replication studies has been to assess the quality of BSRI items in terms of their identification as "masculine," "feminine," or "neutral." Item selection studies have been reported by a number of researchers, including Edwards and Ashworth (1977), Harris (1994), Heerboth and Ramanaiah (1985), Ramanaiah and Hoffman (1984), Walkup and Abbott (1978), and Ward and Sethi (1986). Bem (1981a) incorporated the findings of one such study (Walkup & Abbott, 1978) into her development of the BSRI Short Form in 1978.

Other studies (e.g., Harris, 1994) included additional research questions, such as a comparison of culturally defined masculinity and femininity among ethnic groups. Although Harris (1994) claimed that his results supported the validity of the BSRI items, a closer scrutiny of this study revealed that the sample size was so large ($\underline{N}$ = 3000) that significance was inevitable.

Harris (1994) described his work as "a replication study of item selection for the Bem Sex Role Inventory" (p. 241); however, he asked participants to

evaluate only the items that were ultimately included on the BSRI in terms of being "masculine" or "feminine." This procedure was similar to that followed by Ballard-Reisch and Elton (1992), who assessed a predominantly middle-class, Caucasian, non-college population in the western United States for their interpretations of whether the 60 BSRI items were viewed as "masculine," "feminine," or "neutral." Ballard-Reisch and Elton (1992), however, found that 19 of the 60 items were viewed as neutral, only one was viewed as feminine ("feminine"), and only one was viewed as masculine ("masculine"). Agreement among participants in Ballard-Reisch's (1992) study did not reach the established agreement level (75%) in relation to the remaining 39 items.

As the above review suggests, results of replication studies of item selection procedures used in the development of the BSRI are inconsistent.

Scoring of the BSRI

Bem (1977) modified scoring procedures of the BSRI based upon the criticism of Spence et al. (1975) that her method had no way of discriminating between individuals who scored low on both the Masculine and Feminine scales and those who scored high on both scales. A median-split technique was used instead, resulting in the formation of four distinct groups: feminine, masculine, androgynous, and undifferentiated.

The median-split was used by Spence et al. (1974, 1975) as well in computing scores on their instrument, the Personal Attributes Questionnaire (PAQ). However, Spence and Helmreich (1978) have voiced more concern

than Bem regarding the use of this technique, in that it results in data subject to some statistical distortion. They stressed that, particularly when research questions deal with between-group comparisons, results obtained using this scoring method should be interpreted with caution. An individual's classification is based on reference to others in the sample, and, therefore, test-retest reliability also will be affected (Spence & Helmreich, 1978). Although Bem (1981a) acknowledged that "problematic cases" could result from use of the median-split method, she stated that they "all represent individuals who score near the cutoff point for femininity or masculinity or both. Such cases are an inevitable part of any classification scheme, and they constitute an additional source of 'noise' or error in any research design" (p. 9). However, the observations that (a) this often includes a considerable number of people who score near the cutoff point, and (b) researchers seem to have consistently attached a considerable degree of importance to the classification of their subjects according to the BSRI, as have many subjects themselves, suggest that Bem's perspective here is a serious minimization.

Validity of the BSRI

Any discussion of the validity of the Bem Sex-Role Inventory must begin with a revisiting of the definition of the construct being measured. Indeed, scholars such as Messick (1989) claimed that construct validity superceded all other types of validity. The question of what exactly is being measured by the BSRI has already been raised. In her reply to Spence and Helmreich, Bem

(1981b) argued that they were wrong in their contention that the BSRI has not been shown to tap global masculinity and femininity or gender-schematic processing. Bem's (1981b) explanation that, for sex-typed individuals, the BSRI measures masculinity and femininity, and for non-sex-typed individuals, it does not, is unclear on several counts (Ashmore, 1990). First, the status of cross-sex-typed individuals is a "conceptual and empirical loose end" (Ashmore, 1990, p. 507). Bem (1985) herself has acknowledged that her theory does not address several issues related to cross-sex-typed persons. Secondly, and of even greater concern, is the problem that femininity and masculinity remain inadequately and inconsistently defined in these and other discussions.

Validity studies conducted by Bem and her colleagues (Bem, 1975; Bem & Lenney, 1976; Bem, Martyna, & Watson, 1976) are briefly described in the test manual (Bem, 1981a). Bem sought to verify that the BSRI was able to discriminate between individuals who restricted their behavior in accordance with sex role stereotypes and those who did not. Her primary hypothesis was that a person with a nonandrogynous sex-role classification would demonstrate a more limited range of behavior across a variety of situations (Bem, 1981a). In one such study, subjects were asked to specify which of a series of paired activities they would choose to be photographed performing for pay. Results indicated that sex-typed individuals were significantly more likely than androgynous or cross-sex-typed persons to prefer sex-stereotypical activities (Bem & Lenney, 1976). Bem claimed additional support for the the validity of

the BSRI based upon the results of studies dealing with instrumental and expressive functioning. This research was comprised of four laboratory studies described in two articles (Bem, 1975; Bem, Martyna, & Watson, 1976). The first was designed to measure independence under pressure from a majority to conform. The purpose of the other three was to measure nurturance or emotional responsiveness with a kitten, a baby, and a lonely student, respectively. For men, results were clearer than for women. Specifically, men classified as "feminine" tended not to demonstrate independence and "masculine" men tended not to demonstrate nurturance. While "feminine" women were low in independence and "masculine" women were low in nurturance, as demonstrated in their behavior with the baby and lonely student, "masculine" women did display nurturance with the kitten. Bem (1981a) pointed out that androgynous persons of both sexes demonstrated independence and nurturance as appropriate, depending on the situation. The result that "feminine" women did not display nurturance as predicted is noticeably absent from the discussion offered in the test manual (Bem, 1981a), but is discussed in the report of the actual study (Bem, Martyna, & Watson, 1976). A more thorough description of these and other studies (e.g., Bem, 1981c; Frable & Bem, 1985) can be found in Bem (1985). Bem (1987) contended that "[t]aken as a whole, they provide evidence that sex-typed individuals do, in fact, have a greater readiness than many other individuals to impose a gender-based classification system on reality" (p. 269). In addition, she included in the test

manual (Bem, 1981a) an annotated bibliography of 24 studies that she offered as a reflection of "a growing literature by other investigators that supports the validity of the BSRI by establishing conceptually relevant behavioral correlates" (p. 16).

Despite Bem's research, however, the BSRI's construct validity is questionable (Lippa, 1985; Payne, 1985; Spence, 1984, 1985, 1991). At the very least, results of validity studies conducted by Bem and those conducted by some other researchers are conflictual. Misgivings concerning the construct validity of the BSRI stem from observations that Bem frequently presented contradictory or at least unclear information about what the BSRI is intended to measure. Furthermore, what it actually does measure remains debatable. Spence (1991) concluded that "the BSRI M scale, and to a somewhat lesser degree because of its mixed content, the BSRI F scale" has "construct validity as a measure of desirable instrumentality and expressiveness" (p. 162). Payne (1985) interpreted the "limited validity data that Bem presents" as simply indicative of "some tendency for self-description on the BSRI to agree with overt conduct" (p. 178). Lippa (1985) concluded that "[n]umbers of validation studies suggest that the BSRI femininity and/or masculinity scales are correlated with gender-related behaviors" (p. 177); this conclusion, however, does not adequately address construct validity.

Others have argued that there is sufficient evidence for the construct validity of the BSRI (Brannon, 1978; Lenney, 1991). However, the qualification

that the validity is adequate "when it is used in ways suggested by the theoretical rationale underlying its development" (Lenney, 1991, p. 596) is suspect in light of the problems with the BSRI's theoretical rationale that have been identified.

A thorough investigation of both the content and the process validity of the BSRI conducted by Myers and Gonda (1982) failed to provide support for either type of validity. Not surprisingly, one of Myers and Gonda's major criticisms focused on ambiguities in the definition of masculinity and femininity, which they argued could be interpreted to suggest that in 1982 there was still no meaningful way to operationalize these constructs. With respect to the process validity of the BSRI, Myers and Gonda (1982) argued that "although persons may be aware of stereotypic sex differences, they do not necessarily evaluate themselves in terms of some 'widely known' stereotype when they fill out questionnaires such as the BSRI" (p. 317). The logic of this conclusion is similar to that of arguments presented by Lewin (1984b) and Spence (1985), who repeatedly pointed to the need to allow individuals their own personal definitions of masculinity and femininity. Myers and Gonda (1982) also contended, however, that their findings cannot be used to totally discount that for *some* individuals, the BSRI is a reliable and valid indicator of their sex role orientation" (p. 317).

Reliability of the BSRI

Bem (1981a) reported reliability data based on two samples of Stanford

undergraduates. The first sample, obtained in 1973, included 279 females and 444 males; the second sample, obtained in 1978, included 340 females and 476 males. Coefficient alpha was computed separately for males and females in both samples for the Femininity score, the Masculinity score, and the Femininity-minus-Masculinity score. Analyses were performed separately for the Original BSRI and for the Short Form. Coefficient alphas ranged from .75 to .90, with the Short Form showing higher internal consistency than the Original Form for the Feminine and Femininity-minus-Masculinity scores. It should be noted that the Short Form does not include the items "feminine" or "masculine," nor does it include other items from the Original Form that showed poor item-total correlations with the Masculinity and Femininity scales. Several items deleted from the Femininity scale were ones that were less significantly socially desirable than some of the others (e.g., gullible, childlike, flatterable).

Bem reported test-retest reliabilities over a four-week time span that ranged from .76, for males describing themselves on the masculine items (both Original and Short Forms), to .94, for females describing themselves on the masculine items (Original Form).

Not surprisingly, the Short Form scales correlated strongly (approximately .90) with the corresponding scales of the Original BSRI. With respect to reliability (and validity), the majority of BSRI critics concur that the Short Form can be useful in providing indices of the degree to which individuals describe themselves as "having a global 'instrumental,' 'dominant,' or 'assertive'

disposition and 'expressive' or 'nurturant' tendencies" (Payne, 1985, p. 179).

Lippa (1985) criticized the BSRI manual in that Bem provided little evidence of the discriminant validity of the BSRI scales, aside from a lack of correlation with social desirability as measured by the Marlowe-Crowne Scale. The multitrait-multimethod paradigm (Campbell & Fiske, 1959) provides reliability data in addition to information pertaining to construct validity. Results of a mulitrait-multimethod study conducted by Wong, McCreary, and Duffy (1990) indicated that "while perhaps reliable" (p. 249), the BSRI Masculinity and Femininity scales lacked clear convergent and discriminant validity. Although much less controversy surrounds the BSRI's reliability than its validity, reliability without validity is of questionable value.

Factor Analyses and Dimensionality

Many factor analytic investigations of the BSRI have been conducted (e.g., Antill & Russell, 1982; Gaudreau, 1977; Pedhazur & Tetenbaum, 1979), generally resulting in the conclusion that the scales are not factorially pure. Bem (1979) maintained that this does not suggest inconsistency with the rationale and development of the scales, as societal stereotypes are not necessarily consistent. However, Bem (1981a) also had argued that detailed steps were followed to ensure that BSRI items accurately represent cultural stereotypes. This defense, then, appears to be one more example of vague if not discrepant information presented by Bem.

Factor analyses typically depict two highly correlated instrumentality

factors, one of which can be labelled "dominance" and the other "self-reliance"; an expressiveness factor; and a fourth factor often correlated with biological sex, defined by three BSRI items: "feminine," "masculine," and "athletic" (Lippa, 1985). Consistent with Lippa's review, Pedhazur and Tetenbaum (1979) identified two highly correlated factors related to instrumentality, which they called "assertiveness" and "self-sufficiency"; a factor that tapped expressive traits; and a fourth factor comprised of the items "masculine" and "feminine" in women's self-reports, and defined by "masculine," "feminine," "childlike," and "gullible" in men's self-reports. This fourth factor was bipolar ("childlike" and "gullible" joined with "feminine" in the males' self-reports), as well as being orthogonal to the other factors.

In contrast to other factor analytic studies, Martin and Ramanaiah (1988) found support for the "two-dimensional nature of Bem's Masculinity and Femininity scales in their shortened versions" (p. 348). Martin and Ramanaiah (1988) "support[ed] the conclusion that Bem's short form contains two relatively unidimensional and distinct scales [Masculinity and Femininity] and that the items in each scale share substantial common variance" (p. 348). Martin and Ramanaiah's conclusions, however, are noticeably unsupported by other research.

Several contemporary stalwarts in the field of masculinity and femininity research (e.g., Ashmore, 1990; Marsh & Myers, 1986; Spence, Deaux, & Helmreich, 1985), which necessarily includes BSRI research, have contended

that, in general, the literature would suggest that the collection of attributes and behaviors used to differentiate the sexes are multidimensional. Unfortunately, however, this collection of attributes and behaviors continues to be studied as two unidimensional constructs called "masculinity" and "femininity."

Summary

Issues pertaining to the Bem Sex-Role Inventory, particularly problems related to its theoretical rationale, procedures for item selection, construct validity, and dimensionality, could certainly be interpreted as sufficient evidence to warrant considerable doubt regarding the use of the BSRI in research designed to assess masculinity and femininity. The question we are left with is, Where to go from here?

Theoretical and psychometric problems notwithstanding, the BSRI has served for 20 years as a vehicle for empirical research in masculinity, femininity, and androgyny. It is certainly to the credit of scholars such as Sandra Bem and Janet Spence that we have been challenged to think critically about such constructs. There is much to learn from their contributions. Nevertheless, it is now time to build on their work by ceasing to reinforce the dichotomy between males and females, and beginning to explore the possibilities of the type of society that Bem has come to support.

Androgyny research has responded to the challenge of the bipolarity assumption articulated by Constantinople (1973). But the criticisms of traditional MF research voiced by Constantinople in 1973 and by Lewin in 1984

have remained largely unaddressed, particularly those criticisms that deal with multidimensionality and the identification of meaningful criteria for assessment of masculinity and femininity.

Even though the constructs of masculinity and femininity have proven to be somewhat elusive, and even though the MF tests of the past 60 years appear inadequate, masculinity and femininity are still important and intrinsic concepts to many individuals and to society as a whole, and are therefore worth defining (M. Lewin, personal communication, August 28, 1995). The challenge to do so, however, is overwhelming. It is much easier to simply continue using that with which we are familiar, such as instruments like the BSRI.

Nonetheless, attempts have been made to encourage individual definitions of masculinity and femininity using instruments based on role construct theory (Kelly, 1955). Two role construct repertory tests have been developed: (a) the Sex-Rep (Baldwin, Critelli, Stevens, & Russell, 1986) and (b) the Ravinder Sex Role Salience Reptest (Ravinder, 1987). Neither of these instruments was used in more than a few studies. Both are cumbersome to score. Furthermore, despite the intention to circumvent problems related to stereotyping, an invitation to stereotype remains implicit. For example, respondents are asked to describe ways in which other people are "masculine," or "feminine," which may be very different from the respondent's definition of "masculinity" or "femininity" in relation to self.

Hence, the challenges of responding to the extant criticisms of MF

involving multidimensionality and the establishment of meaningful criteria for assessment of masculinity and femininity have remained. An examination of the gender identity construct may provide the key to untangling these issues.

Gender Identity

Gender identity has been described as "a basic, existential conviction that one is male or female" (Spence & Sawin, 1985, p. 59); a secure sense or conviction of one's own maleness or femaleness (cf. Green, 1974; Money, 1994); and the "individual's awareness of and satisfaction with being a male or female" (Pleck, 1984, p. 220). As such, gender identity refers to one's subjective feelings of maleness or femaleness (Basow, 1992; Golombok & Fivush, 1994) and a sense of confidence in and comfort with being either male or female (Lewin, 1984b). In essence, gender identity is an individual's concept of himself or herself as male or female (Golombok & Fivush, 1994).

Rather than equating masculinity and femininity with stereotypical gender traits and roles, masculinity and femininity can be reconceptualized in terms of the gender identity construct, and, thus, as part of one's self-concept. Spence (1985) proposed that "masculinity and femininity, as they refer to an individual's self-concept, be retained and conceptualized as gender identity: a basic phenomenological sense of one's maleness and femaleness that parallels awareness and acceptance of one's biological sex and is established early in life" (p. 91). Lewin (1984b) suggested that masculinity and femininity be conceptualized as "the gender-relevant aspects of a person's self-concept," thus

allowing for "individual variation in the specific content of the self-image as related to gender" (p. 200).

Part of the challenge to researchers seeking to measure masculinity and femininity via gender identity is that the terms "gender identity" and "gender role identity" are frequently used interchangeably. Gender role identity, as opposed to gender identity, refers to the degree to which a person identifies with societal, not personal, definitions of masculinity and femininity (cf. Basow, 1992). Gender role identity is the construct at the core of instruments such as the Bem Sex-Role Inventory and the Personal Attributes Questionnaire. However, non-adherence of males and females to societally prescribed gender roles does not necessarily imply uncertainty regarding one's gender identity (Golombok & Fivush, 1994). In addition, not only has "gender role identity" been confused with "gender identity," it also has been indiscriminately used as a synonym for masculinity and femininity. However, as Spence (1985) suggested, individuals may create their own standards or "calculus" for self-assessing maleness or femaleness. For example, Tangri (1972) found that women whose career aspirations were traditionally "masculine" did not consider themselves to be "masculine"; in contrast, they defined their femininity in a variety of other ways. It is clear that, in gender research, caution should be exercised when choosing words to describe the constructs being studied (Sherif, 1982; Unger & Crawford, 1993).

To maintain a clear focus on gender identity as a construct by which we

can better understand masculinity and femininity is no small task. Constantinople (1973) pointed out that the relationship between a theoretical definition and the measures of masculinity and femininity was further complicated by indiscriminate usage of the terms "sex role adoption" and "sex role preference." She argued that the two latter terms are far from synonymous in that an individual may prefer to have characteristics associated with one sex, but, in actuality, tend to exhibit traits primarily associated with the other sex. Constantinople (1973) observed that, in the empirical literature,

> [t]his extrapolation from either preference or adoption to M or F is made more often than not. Conceptually, however, neither [preference nor adoption] would seem to be an adequate indicator of M-F; rather, there seems to be some notion of *identity* that should be included when making a statement about an individual's masculinity or femininity. (p. 391)

Even prior to the work of Spence (1978) and Lewin (1984b), then, we can see that other earlier scholars (e.g., Constantinople, 1973) had argued that masculinity and femininity have much more to do with gender identity, as described above, than with gender role identity. To describe the nature of the individual's self-concept as he or she relates it to masculinity or femininity would indeed be a more fruitful approach to understanding human behavior than counting the number of ways in which an individual resembles the typical member of his or her own sex (Spence, 1978).

In her description of gender identity, Spence (1985) argued that, as long

as individuals see themselves as possessing attributes that they perceive as relevant to their own gender, their masculinity or femininity is taken for granted and is infrequently a topic of conscious concern or reflection. This might help to explain why people, regardless of educational background, etc., have difficulty articulating their personal meanings of masculinity or femininity. In the Spence and Sawin (1985) study, for example, individuals questioned regarding their senses of masculinity and femininity responded with such statements as "It's something I've never had to think about," "I've never had a problem with masculinity," and "I'm a woman and always have been" (Spence, 1985, p. 83).

Spence (1985) suggested that "[w]hat constitutes an adequate amount of gender-relevant qualities for a given individual is determined by a complex calculus operating below the level of conscious awareness" (p. 83). Individuals' definitions of masculinity and femininity not only vary from person to person, but also may differ when individuals are assessing themselves as opposed to others (Spence, 1985). Spence contended that, on some level, people strive to keep their sense of masculinity or femininity intact, using characteristics they attach to gender and that they possess to confirm their gender identity. Spence (1985) and Money (1994) argued that most men and women appear to be more secure than not in their gender identity most of the time. However, developmental tasks (e.g., adolescence) or life events (e.g., divorce) may result in stresses that cause people to periodically doubt their masculinity or femininity and struggle to reaffirm it (Spence, 1985).

Spence (1985) argued that issues more central to one's identity than developmental or life events (e.g., identity as a lesbian or gay man) can be related in various ways to one's sense of femininity or masculinity. For example, the belief held by a number of heterosexual individuals that lesbians and gay men cannot be "real" women and men reflects the importance that some heterosexual men and women place on sexual orientation in evaluating their own and others' masculinity or femininity. In contrast, however, many lesbians and gay men clearly define their femininity and masculinity totally separate from sexual orientation and feel confident as females and males (Spence, 1985).

From Spence's discussion we can extrapolate that a sense of confidence in relation to gender, or a sense of adequacy as a man or woman, is tied to one's gender identity, and, hence, to one's personal sense of masculinity or femininity. Furthermore, the work of scholars such as Lewin (1984b) is consistent with the argument that an individual's sense of himself or herself is the critical dimension in a discussion of masculinity and femininity measurement. Lewin argued that one's "perceived self-image will be more compelling than [one's] perceived traits or interests" (p. 200).

Gender identity, then, encompasses gender self-concept. Like self-concept in general, gender self-concept is multifaceted. Confidence in self has been identified as one important aspect of self-concept (Basow, 1992; Hattie, 1992), and, consequently, of gender self-concept. This argument is consistent

with Lewin's (1984b) assertion that MF tests should assess gender self-confidence. Like Spence, Lewin argued that the focus should be on measuring "individuals' beliefs that they are, or are not, living up to various aspects of their personal gender-relevant self-concepts. Do they feel competent as members of their own sex? Are they meeting their own standards for masculinity and femininity?" (p. 200). It would seem most appropriate, then, to investigate gender self-confidence as one aspect of gender self-concept, and, hence, of gender identity, as a means toward better understanding masculinity and femininity.

One might ask why, over a decade since Lewin (1984b) articulated these concepts, such an investigation has yet to be attempted. One's subjective feelings of maleness or femaleness have been largely ignored, while the stereotypically "feminine" and "masculine" conceptions of femininity and masculinity have continued to provide the standards by which we assess these constructs. Scholars of women's studies (e.g., Cook, 1993; Worell & Remer, 1992) and men's studies (Kimmel & Messner, 1995; Rabinowitz & Cochran, 1994) have repeatedly acknowledged the importance to women and men of developing, using, and integrating one's expressive (traditionally "feminine") and one's instrumental (traditionally "masculine") identities. Nonetheless, there is a tendency toward what Spence (1978) called the "path of least resistance" (p.117), where either expressivity (traditional "femininity") or instrumentality (traditional "masculinity") is emphasized at the expense of the other, where

females and males alike feel compelled to support the stereotypic notions of what it means to be women and men. Hence, perceptions of the characteristics of women and men, when considered as a totality as opposed to the particular, become "exaggerated and polarized" (Spence, 1978, p. 117). Our reluctance to revise stereotypical notions of masculinity and femininity makes challenging these traditional definitions difficult. Almost 20 years ago, Spence (1978) noted that, "[p]articularly among those whose sense of 'personhood' is not secure or well developed and who rely on their correspondence to traditional standards of behavior for their definition of self, attacking these standards may be attacking one of the central aspects of their self-identity" (p. 117). If some of this resistance can be better understood, then perhaps we can overcome it and move onward.

Conclusion

This chapter has addressed the literature related to five key areas that are central to this study. First, a history of masculinity and femininity measurement through 1970 was presented. Assumptions associated with traditional masculinity and femininity measurement were explored. Secondly, a new era in the measurement of masculinity and femininity, highlighted by the work of Sandra Bem, Janet Spence, and Meda Rebecca, was described. Concepts of masculinity and femininity were revised during this period; however, a dichotomy between the two constructs persisted. Next, the androgyny bandwagon was discussed, including an overview of the research

that ensued from this revolution in thought. The fourth section of this chapter detailed major critiques of Sandra Bem's monumental work, the Bem Sex-Role Inventory. The final section considered the gender identity construct, and how it may provide the key to a more meaningful conceptualization of masculinity and femininity through the assessment of gender self-confidence.

Given the problems with current measurement of masculinity and femininity that derive from inadequate definitions and faulty assumptions, and, given the potential of other constructs (e.g., gender self-confidence) as vehicles toward more successful assessment of masculinity and femininity, the purpose of the proposed study is to re-examine the androgyny construct and its measurement using this information. A review of the related literature supports the need to further explore the currency of the Bem Sex-Role Inventory (BSRI) as representative of perceptions held by college undergraduates of "feminine" and "masculine" characteristics. The literature further suggests that the concept of gender identity as a representation of masculinity and femininity merits investigation. Gender self-confidence has been identified as one aspect of gender identity. Preliminary steps to assess gender identity will be taken by examining gender self-confidence in this regard. Such steps include, first, the development and refinement of an instrument designed to assess gender self-confidence, and, second, an exploration of the relationship between participants' levels of gender self-confidence and information obtained using the BSRI. The methodology utilized to conduct these steps is the subject of the following

chapter.

CHAPTER III

METHODOLOGY

A review of the related literature supports the contention that the measurement of masculinity and femininity, as it currently exists, is inadequate. The literature also suggests that masculinity and femininity might be better represented using the concept of gender identity, and that the assessment of gender self-confidence, as one component of gender identity, is warranted as a step toward a clearer understanding of these constructs. This chapter presents the design and methodology for the study. Included are research questions and hypotheses; descriptions of participants, instruments, and procedures; and statistical analyses of the data.

Research Questions and Hypotheses

This study had three goals: (a) to examine the current viability of the BSRI as a research tool by assessing whether its "masculine" and "feminine" items represent current perceptions of masculinity and femininity among college undergraduates, (b) to explore the constructs of masculinity and femininity as representations of gender identity as opposed to umbrellas for stereotypically viewed personality traits, and (c) to assess gender self-confidence as a component of gender identity.

Specifically, this study addressed the following research questions:

1. How do college undergraduates currently describe themselves using the BSRI?

2. Do the "masculine" and "feminine" items on the BSRI represent current perceptions of masculinity and femininity among college undergraduates?

3. What are the major dimensions of college undergraduates' self-reported levels of gender self-confidence?

4. What is the relationship between college undergraduates' levels of gender self-confidence and their self-descriptions according to BSRI classifications (i.e., masculine, feminine, androgynous, undifferentiated)?

5. What is the relationship between college undergraduates' levels of gender self-confidence and their evaluations of BSRI items as masculine and feminine?

Research Hypotheses

The following research hypotheses were formulated based upon the research questions:

1. There will be a significant difference between the percentage of college undergraduates classified as feminine, masculine, androgynous, and undifferentiated in 1978, using a median-split classification system based upon Bem's 1978 data, and the percentage of college undergraduates classified as feminine, masculine, androgynous, and undifferentiated in 1995, using a median-split classification system based upon the data from this sample.

2a. There will be no significant agreement among college undergraduates supporting the "masculinity" of the items that comprise the BSRI Masculine scale.

2b. There will be no significant agreement among college undergraduates supporting the "femininity" of the items that comprise the BSRI Feminine scale.

3. There will be two factors that define college undergraduates' levels of gender self-confidence, identified as (a) gender self-definition and (b) gender self-acceptance.

4. There will be no significant relationship between college undergraduates' levels of gender self-confidence, as measured by the Hoffman Gender Scale (Hoffman, 1996), and their self-descriptions according to BSRI classifications.

5. There will be a significant positive relationship between college undergraduates' levels of gender self-confidence, as measured by the Hoffman Gender Scale (Hoffman, 1996), and their neutral evaluations of BSRI "masculine" and "feminine" items, as measured by an evaluation score for each respondent.

Participants

Participants in this study were 371 undergraduate students (273 females and 98 males) in attendance at an ethnically diverse, medium-sized, public university located in the southern United States. Undergraduate students were

selected as the sample for this study in order to maximize comparison with the norms established by Bem in 1978. To enhance the likelihood of a balance of males and females, as well as a representative sample of the undergraduate college population, the researcher recruited participants from courses in several departments within the university. These included the departments of: (a) Public Health Education, in the School of Health and Human Performance, (b) Counseling and Educational Development, in the School of Education, and (c) Human Development and Family Studies, in the School of Human Environmental Sciences. Student-athletes served by the university's Academic Enhancement Program also were recruited.

A power analysis determined that a sample size of 198 was sufficient for medium effects and to ensure a power of .80 (Cohen, 1977). Thus, the sample (N = 371) was more than adequate to meet these standards.

Instrumentation

Each participant in the study was asked to perform the following tasks: (a) complete the Bem Sex-Role Inventory (BSRI) as a self-report measure, (b) respond to each of the 60 items that comprise the BSRI by indicating whether they are perceived as "feminine," "masculine," or "neutral," and (c) complete the Hoffman Gender Scale (HGS) as a self-report measure.

Bem Sex-Role Inventory (BSRI)

The Bem Sex-Role Inventory (BSRI; Bem, 1974), including its theoretical rationale, item selection procedures, score interpretation, construct validity,

reliability, and dimensionality, as well as critiques of all of the above, were discussed in detail in Chapter II (pp. 29-52). As a consequence, only an overview of the BSRI is presented here.

The BSRI was developed to facilitate empirical research on psychological androgyny. Unlike developers of previous MF measures, Bem (1981a) designed her instrument to measure masculinity and femininity as independent dimensions. While described as a measure of masculinity and femininity, the BSRI also was purported to assess gender schematicity, or the degree to which an individual tended to encode and organize information, including information about the self, in terms of cultural definitions of maleness and femaleness (Bem, 1981a). Gender schema theory (Bem, 1981b, 1981c), which included assumptions about the cognitive processing and the motivational dynamics of sex-typed versus androgynous individuals, provided the theoretical rationale for the development of the BSRI, although the theory was not fully developed nor named at the time that the BSRI was developed.

The BSRI consists of 60 personality characteristics on which respondents are asked to rate themselves using a 7-point Likert scale (1 = Never or almost never true; 7 = Always or almost always true). Twenty of the characteristics are stereotypically feminine (e.g., affectionate, gentle, sympathetic), 20 are stereotypically masculine (e.g., independent, forceful, dominant), and 20 are considered filler items by virtue of their gender neutrality (e.g., conscientious, conceited, truthful). Unlike the "feminine" items and the

"masculine" items, which were all identified as socially desirable for their respective sex, 10 of the "gender-neutral" items were identified as desirable for both sexes (e.g., adaptable, sincere) and the other 10 as undesirable for both sexes (e.g., inefficient, jealous). These 20 items were used to comprise a measure of Social Desirability in response.

Scoring of the BSRI involves the use of a median-split classification system whereby four distinct quadrants are formed. Individuals are classified as "androgynous" if they rate higher than the median on both dimensions (masculinity and femininity), "undifferentiated" if they rate lower than the median on both dimensions, "feminine" if they rate higher than the median on femininity but lower than the median on masculinity, and "masculine" if the reverse is true.

In the construction of the BSRI, Bem (1974) used independent t-tests to ascertain whether each of the 400 items in her original pool of adjectives was significantly more desirable for a man than for a woman (her criterion for qualifying as "masculine" based on ratings of 50 undergraduate students who served as judges), or for a woman than a man (her criterion for qualifying as "feminine" based on ratings of a second group of 50 undergraduates), or neither.

Validity of the BSRI is dependent upon the definition of the construct being measured. The BSRI has been said to assess masculinity and femininity; it also has been said to assess gender schematicity (Bem, 1981a). Bem and her colleagues (Bem, 1975; Bem & Lenney, 1976; Bem, Martyna, & Watson,

1976) concluded that validity was established by verifying that the BSRI could discriminate between individuals who restricted their behavior in accordance with sex role stereotypes and those who did not.

Bem (1981a) reported reliability data based on two samples of Stanford undergraduates. The first sample, obtained in 1973, included 279 females and 444 males; the second sample, obtained in 1978, included 340 females and 476 males. Coefficient alpha was computed separately for males and females in both samples for the Femininity score, the Masculinity score, and the Femininity-minus-Masculinity score. Coefficient alphas for the Original BSRI ranged from .75 to .87.

Bem (1981a) reported test-retest reliabilities over a four-week time span that ranged from .76, for males describing themselves on the masculine items, to .94, for females describing themselves on the masculine items (Original Form).

Hoffman Gender Scale (HGS)

The development of an instrument designed to assess gender self-confidence formed the basis for the pilot study, conducted in August and September 1995. Gender self-confidence is defined as that aspect of gender identity related to one's self-assuredness about being male or female, and is measured by assessing "individuals' beliefs that they are, or are not, living up to various aspects of their personal gender-relevant self-concepts" (Lewin, 1984, p. 200).

The construct was clarified using a content analysis, review of relevant literature, and expert judgment. Content analysis consists of posing open-ended questions to individuals about the construct of interest, and then sorting their responses into topical categories (Crocker & Algina, 1986). Expert judgment, which involves the test developer soliciting input from individuals who have studied the construct of interest (Crocker & Algina, 1986), was obtained through personal communication with Miriam Lewin (August 28, 1995). Lewin concurred that the approach utilized by the test developer was appropriate. An item review was conducted in which qualified colleagues and members of a graduate class in test development were asked to informally assess the items for wording, accuracy, ambiguity, and other technical flaws.

Description of the instrument. The resulting instrument consisted of 20 statements to which respondents were asked to indicate their level of agreement according to a 6-point Likert format, ranging from 1 (Strongly Agree) to 6 (Strongly Disagree). Labels also were assigned to the intermediary categories (2 = Moderately Agree; 3 = Tend to Agree; 4 = Somewhat Agree; 5 = Disagree). These labels were selected in an attempt to increase reliable variance (see Klockars & Yamagishi, 1988; Lam & Stevens, 1994). Two forms of the instrument were used in order to improve readability based on sex of respondent: Form A was worded for a female audience and Form B was written for males. Other than substituting "male" for "female" and "masculinity" for "femininity," all items were identical between the two forms. Following the

20 items was the question: "What do *you* mean by femininity (or masculinity)?" This question was included as a means to explore the idiosyncratic meanings of the constructs that the respondents were self-assessing. The instrument was described as the Hoffman Gender Scale (HGS; Hoffman, 1996) (Table 1) in order to reduce the emphasis on the self-confidence component.

Participants. The HGS was administered by the researcher to students enrolled in seven undergraduate classes in the counseling department of a moderately-sized university in the southeastern United States. The classes included five sections of Career and Life Planning and two sections of Helping Skills. Class time was used for administration. Although participation was voluntary, no students declined. Respondents included 92 females and 54 males.

Descriptive and item-total statistics. For ease of statistical analysis, each of the 20 items was described using a word or an abbreviation (see Table 2). Means, standard deviations, item-total correlations, and factor loadings are provided in Table 2. The results of initial descriptive analyses suggested that the two anchors Moderately Agree and Tend to Agree may not have been well chosen so as to ensure reliable variance, and might need to be changed in future versions.

It should be noted that because items 5 and 10 were worded negatively, they required reverse-scoring (i.e., 1 = 6; 2, 3, and 4 = 5; 5 = 3; and 6 = 1).

Coefficient alpha was computed to assess internal consistency of the

Table 1

Hoffman Gender Scale (Form A)

Please indicate your level of agreement with each of the following statements by rating it a "1," "2," "3," "4," "5," or "6" as follows:

1	2	3	4	5	6
Strongly Agree	**Moderately Agree**	**Tend to Agree**	**Somewhat Agree**	**Disagree**	**Strongly Disagree**

1.	When I think of myself as a female, I really feel good.	____
2.	I am confident in my femininity.	____
3.	I feel competent as a member of my sex.	____
4.	My self-image is feminine.	____
5.	I have doubts about my femininity.	____
6.	I meet my personal standards for femininity.	____
7.	My perception of myself is positively associated with my biological sex.	____
8.	I am proud to be a female.	____
9.	I am secure in my femininity.	____
10.	Being a female makes me uneasy.	____
11.	I think of myself as feminine.	____
12.	I define myself largely in terms of my femininity.	____
13.	I accept myself as a member of my biological sex.	____
14.	My identity is strongly tied to my femininity.	____
15.	I have a high regard for myself as a female.	____
16.	I am happy with myself as a female.	____
17.	I am very comfortable being a female.	____
18.	My sense of myself as a female is positive.	____
19.	Being a female contributes a great deal to my sense of confidence.	____
20.	Femininity is an important aspect of my self-concept.	____

What do *you* mean by femininity?

_____

Table 2

Descriptive Statistics by Sex of Respondent and Factor Loadings

Item #	Item	Females			Males			
		Mean	SD	Item-Total Correlations	Mean	SD	Item-Total Correlations	Factor Loadings
1.	FEELGOOD	1.54	.93	.70	1.56	.97	.79	.72
2.	CONFID	1.49	.87	.81	1.44	.72	.68	.78
3.	COMPET	1.46	.90	.44	1.41	.80	.63	.44
4.	IMAGE	1.54	.94	.78	1.76	1.12	.76	.74
5.	DOUBTS	1.82	1.38	.46	2.06	1.57	.45	.44
6.	STAND	1.55	.95	.77	1.44	.84	.36	.70
7.	BIOLOGY	1.80	1.19	.48	1.74	1.17	.61	.45
8.	PROUD	1.23	.73	.68	1.41	.90	.78	.73
9.	SECURE	1.38	.80	.79	1.33	.67	.52	.72
10.	UNEASY	1.71	1.28	.79	1.70	1.18	.76	.50
11.	MEFEMALE	1.61	.91	.74	1.60	1.14	.88	.78
12.	DEFINE	2.42	1.32	.75	2.30	1.48	.81	.69
13.	ACCEPT	1.23	.67	.55	1.17	.54	.46	.46
14.	IDENTITY	2.19	1.28	.67	2.30	1.43	.82	.65
15.	REGARD	1.50	.98	.73	1.57	1.04	.80	.79
16.	HAPPY	1.36	.91	.68	1.19	.55	.60	.79
17.	COMFORT	1.24	.73	.73	1.20	.60	.66	.87
18.	SENSE	1.34	.82	.69	1.26	.68	.72	.86
19.	CONTRIB	2.14	1.16	.55	2.19	1.47	.70	.59
20.	SELFCONP	1.95	.98	.67	2.11	1.37	.79	.66

N = 146 (Females = 92; Males = 54)

instrument for both females and males. The sample size (N = 146) was more than sufficient to accurately assess this type of reliability. The alpha coefficient for the female sample was 0.94; for males, it also was 0.94.

A MANOVA was calculated to assess a possible difference between item responses of males and females. There were no overall differences between males and females across the 20 items, $F(20, 125) = .96$, $p = .515$.

Factor Analyses. A maximum likelihood factor analysis clearly indicated the presence of one factor, which explained 50% of the total variance. Taking standard test development procedures and subsequent analyses into consideration, it was apparent that gender self-confidence was the underlying factor.

Further scrutiny of this single factor revealed that although one factor described as gender self-confidence was evident, a continuum ranging from gender self-acceptance to gender self-definition appeared to exist. As indicated by Figure 1, the top five items seemed to represent a construct that would describe gender self-definition; the bottom seven described gender self-acceptance.

Conclusions and Recommendations. The pilot study resulted in support for the use of the HGS with possible modifications to ensure a more sophisticated instrument. An analysis of the pilot study led to the following recommendations for revision of the scale:

1. Modify the existing scale to increase discrimination between

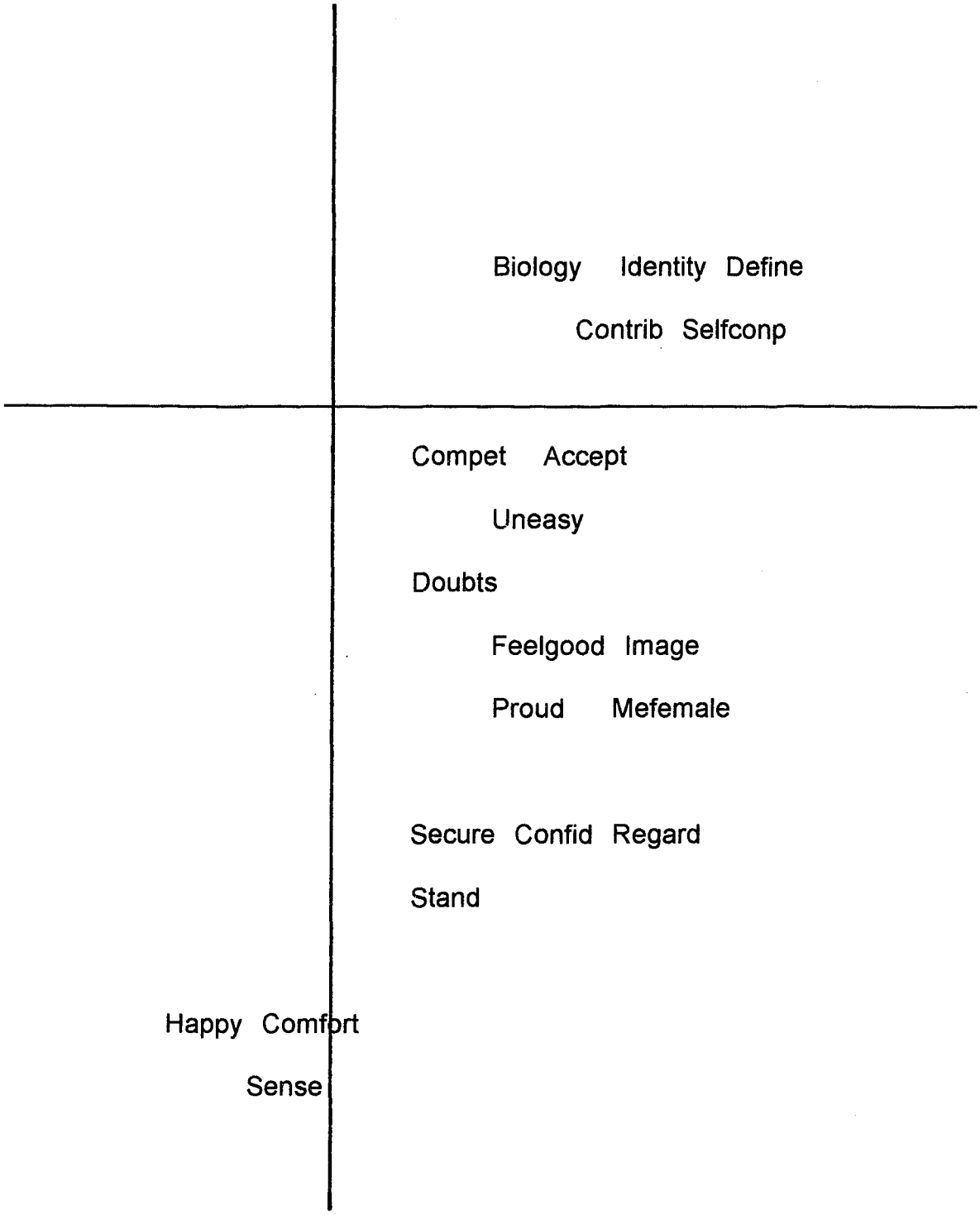


Figure 1. Plot of a Factor Analysis of Hoffman Gender Scale Items

certain categories (e.g., replace Moderately Agree as the second anchor with Agree).

2. Eliminate test items that converge around the middle of the continuum (Compet, Accept, Uneasy, Doubts, Feelgood, Image, Proud, Mefemale) (see Figure 1).

3. Add two test items to the upper end of the continuum to strengthen the "definition" construct of the gender self-confidence factor and to provide an equal number of items at both ends of the continuum.

These three recommendations resulted in the revised version of the Hoffman Gender Scale (See Table 3). The test items that were added were: (1) "When I am asked to describe myself, being female is one of the first things I think of" and (9) "Being a female is a critical part of how I view myself."

To ensure that female and male respondents completed the appropriate form of the instrument, directions to this effect were clarified at the top of the instrument. Forms A (Female) and Form B (Male) of the revised Hoffman Gender Scale, used in the study, are found in Appendices A and B, respectively.

Procedures

The researcher contacted several faculty members and course instructors in the various departments of the university at which the study will be conducted. This was done to obtain information related to class sizes and gender distribution among the classes, as well as to ensure that any

Table 3

Hoffman Gender Scale (Form A)

PLEASE NOTE: Complete Form A if you are a female. Complete Form B (reverse side) if you are a male.

Please indicate your level of agreement with each of the following statements by rating it a "1," "2," "3," "4," "5," or "6" as follows:

1	2	3	4	5	6
Strongly Agree	**Agree**	**Tend to Agree**	**Somewhat Agree**	**Disagree**	**Strongly Disagree**

1. When I am asked to describe myself, being female is one of the first things I think of. ____
2. I am confident in my femininity. ____
3. I meet my personal standards for femininity. ____
4. My perception of myself is positively associated with my biological sex. ____
5. I am secure in my femininity. ____
6. I define myself largely in terms of my femininity. ____
7. My identity is strongly tied to my femininity. ____
8. I have a high regard for myself as a female. ____
9. Being a female is a critical part of how I view myself. ____
10. I am happy with myself as a female. ____
11. I am very comfortable being a female. ____
12. Femininity is an important aspect of my self-concept. ____
13. My sense of myself as a female is positive. ____
14. Being a female contributes a great deal to my sense of confidence. ____

What do *you* mean by femininity?

_____

_____

departmental procedures for entry into classes were noted and followed. Data collection was conducted in November 1995.

The assessment package consisted of the Bem Sex-Role Inventory (BSRI), a listing of all BSRI items to evaluate as "feminine," "masculine," or "neutral," and the revised Hoffman Gender Scale (HGS) (Form A, Appendix A; Form B, Appendix B). All but one administration of the assessment package were conducted by the researcher in order to enhance standardization of procedures. The administration that was not conducted by the researcher was performed by a colleague who was trained by the researcher in the administrative procedures. Instructions to all participants were read by the test administrator (see Appendix C).

All participants were instructed first to complete the BSRI as a self-assessment. In the first six of the 12 test administrations, participants then were asked to go through a listing of the BSRI items and rate each of the 60 items as "feminine," "masculine," or "neutral," using their individual perceptions of these terms as a guide. The third and final task for these participants was to complete the HGS as a self-assessment. (See Appendix C for explicit instructions.) In the last six of the 12 test administrations, the order of task two and task three was reversed. The total time required to complete the assessment packet was approximately 25 minutes.

Data Analyses

Statistical analytic procedures were selected to address each of the

research questions and hypotheses as appropriate. To address hypothesis 1, participants' self-descriptions on the BSRI was scored using two methods: (a) a median-split classification system based upon scores derived from this sample and (b) the hybrid method for classifying individuals that uses both the Femininity-minus-Masculinity score and the median split as bases of classification (Bem, 1981a). Hypotheses 2a and 2b involved calculation of an index of neutrality for each BSRI item based upon participants' assessments of the masculinity, femininity, or neutrality of each item. An established agreement level of 75% (Ballard-Reisch & Elton, 1992; Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972) was used.

As in the pilot study, descriptive statistics were calculated for each of the items of the HGS to address hypothesis 3. Means, standard deviations, and item-total correlations were calculated separately for males and females. Internal reliability was determined by calculating the alpha coefficient for both genders. MANOVA was used to assess possible differences between responses of males and females. Factor analyses also were conducted to assess the dimensionality of gender self-confidence.

Hypothesis 4 was addressed by using MANOVA to examine the differences in means of the gender self-confidence scales when participants were divided into the four classification categories used by Bem (1981a).

Finally, hypothesis 5 was addressed by calculating the Pearson product-moment correlations between each of the gender self-confidence factor scores

and an evaluation score that was established for each participant based upon his or her evaluations of BSRI items as "masculine," "feminine," or "neutral."

CHAPTER IV

RESULTS AND DISCUSSION

In this chapter, the results of the study are presented. First, demographics of the sample are described and examined for possible differences among respondents in terms of age, year in school, and ethnicity. Variation in the order in which instruments were administered also is examined for possible effect. Then, for each of the five research hypotheses stated in the previous chapter, the corresponding data analyses are described and the findings presented. Differences related to sex of respondent are discussed within the context of each research hypothesis.

Influence of Demographic Variables and Order of Instruments

As a preliminary check, the data were examined to determine if age, year in school, race/ethnicity, or the order in which the instruments were administered was related to participants' responses. Detailed information pertaining to the relationship between these variables and scores on the various instruments is provided in this section.

Ninety-eight males and 273 females participated in this study. As indicated above, comparisons of the data according to sex of respondent are described in greater detail in the discussions of the data analyses that correspond to each research hypothesis.

In order to conduct data analyses in terms of race/ethnicity, participants were classified as majority (White/Caucasian) or minority (African American, Hispanic, Native American, Asian, Other) status. This classification was used because the numbers of Hispanic, Native American, Asian, and Other respondents were not sufficient to analyze separately. Of the 335 respondents who indicated their race and year in school, 244 were majority and 91 were minority status. Classified according to year in school, there were 132 freshmen, 84 sophomores, 70 juniors, and 49 seniors. MANOVAS indicated that neither race nor year in school was related to participants' self-descriptions on the BSRI nor to their evaluations of BSRI items as "masculine," "feminine," or "neutral" (see Table 4). Whereas year in school was not related to HGS subscale scores, there were differences in the HGS scores related to race. More specifically, the MANOVA results indicated that respondents in the ethnic minority groups had stronger Gender Self-acceptance scores than majority (Caucasian) respondents (mean for minority respondents = 1.38; mean for majority respondents = 1.63, with lower scores indicating higher gender self-acceptance [F (1, 327) = 11.20, $p < .001$]).

Age of respondents ranged from 17 to 46 (N = 328, M = 20.45, SD = 4.12, median = 19). Table 5 provides information pertaining to the number and percentage of participants of each age. Correlations between age of respondent and scores on the various instruments are provided in Table 6. As indicated in Table 6, no significant correlation was found between age of

Table 4

MANOVA Effects of Race and Year in School on BSRI Scores, HGS Scores, and Evaluation of BSRI Items as "Masculine," "Feminine," or "Neutral"

BSRI	df	F	p
Race	2, 326	.46	.630
Year in School	6, 652	.93	.474
Race x Year in School	6, 652	1.49	.178

Evaluation of BSRI Items	df	F	p
Race	1, 322	3.19	.075
Year in School	3, 322	1.55	.202
Race x Year in School	3, 322	1.53	.206

HGS	df	F	p
Race	2, 326	6.17	.002**
Year in School	6, 652	1.13	.341
Race x Year in School	6, 652	.50	.809

**Note: Gender Self-acceptance, F (1, 327) = 11.20, $p < .001$
Gender Self-definition, F (1, 327) = .15, $p = .695$

Table 5

Frequency and Percentage of Respondents By Age

Age	Frequency	Percentage
17	2	.6
18	101	30.8
19	78	23.8
20	47	14.3
21	40	12.2
22	20	6.1
23	6	1.8
24	11	3.4
25	4	1.2
26	3	.9
28	2	.6
29	1	.3
30	2	.6
33	1	.3
34	1	.3
36	1	.3
37	2	.6
39	2	.6
40	1	.3
43	1	.3
43	1	.3
45	1	.3
46	1	.3

Table 6

Correlations Between Age of Respondent and BSRI Scale Scores, HGS Scores, and Evaluation of BSRI Items as "Masculine," "Feminine," or "Neutral"

BSRI (Original Form)	df	r	p
Masculine Scale	325	.02	> .05
Feminine Scale	325	-.01	> .05

BSRI (Short Form)	df	r	p
Masculine Scale	325	.02	> .05
Feminine Scale	325	-.04	> .05

Evaluation of BSRI Items	df	r	p
Masculine	325	-.06	> .05
Feminine	325	-.07	> .05
Neutral	325	.07	> .05

HGS	df	r	p
Gender Self-definition	325	.15**	< .001
Gender Self-acceptance	325	.24**	< .001

respondent and BSRI scale scores or participants' evaluation scores of BSRI items. There were significant positive correlations between age of respondent and HGS Self-definition scores (r = .15, df = 325, p < .001), and age of respondent and HGS Self-acceptance scores (r = .24, df = 325, p < .001), suggesting that the younger participants were more gender self-confident than the older participants.

As indicated in Chapter Three, all participants in the study completed the BSRI as a self-description prior to their evaluation of BSRI items as feminine, masculine, or neutral, and prior to completion of the HGS. Fifty-eight percent of the respondents (N = 215) completed the evaluation task as the next assignment (prior to completing the HGS). Forty-two percent of the respondents (N = 156) completed the HGS as the second assignment (prior to performing the evaluation task). The results of two MANOVAS suggested no differences in the performance of respondents on the HGS scales related to the two orders of administration (Mult F = 2.45, df = 2, 268, p = .09), or on their evaluation scores related to order (Mult F = 3.48, df = 1, 369, p = .063).

In general, then, none of these demographic variables nor the order of the instruments was found to be a confounding variable that required further analysis by subgroups. Thus, analyses of the data according to each of the five research hypotheses were performed.

Analyses for Research Hypotheses

Research Hypothesis One

There will be a significant difference between the percentage of college undergraduates classified as feminine, masculine, androgynous, and undifferentiated in 1978, using a median-split classification system based upon Bem's 1978 data, and the percentage of college undergraduates classified as feminine, masculine, androgynous, and undifferentiated in 1995, using a median-split classification system based upon the data from this sample.

To address research hypothesis one, participants' self-descriptions on the BSRI were scored using two methods: (a) a median split classification system based upon scores derived from this sample, and (b) the hybrid method for classifying individuals that uses both the Femininity-minus-Masculinity score and the median split as bases of classification (Bem, 1981a).

Before comparing the similarities or differences between Bem's 1978 data and the data derived from the current sample, the reliability for the Masculine and Feminine scales was examined for both sexes. In addition, construct validity was examined using factor analyses.

For females in this study, the estimate of reliability (coefficient alpha) on the Masculine scale was .84; and, on the Feminine scale, it also was .84. For males, coefficient alpha was .85 on the Masculine scale and .80 on the Feminine scale. These estimates provide considerable confidence in the scales

as measures of attributes, although these attributes may be other than masculinity and femininity.

A factor analysis was calculated so that the structure derived from this data could be compared to the findings reported by Bem as well as those described in other factor analytic investigations of the BSRI (e.g., Antill & Russell, 1982, Gaudreau, 1977; Martin & Ramanaiah, 1988; Pedhazur & Tetenbaum, 1979). Table 7 presents the factor loadings from an oblique rotation (oblimin), with the expected loadings in bold. In most cases, the loadings were as anticipated by the scoring suggested by Bem (1981). Consistent with the majority of BSRI factor analytic studies, the exceptions were items found in the Original (long) form of the BSRI. In this study, these included the following "feminine" items, which did not load on either factor: yielding (32), shy (38), flatterable (41), soft-spoken (47), gullible (50), and childlike (53). There were fewer exceptions among the "masculine" items, as most loaded as expected; exceptions included athletic (34) and analytical (37). The correlations between the two factors were close to zero ($\underline{r}$ = .07 for Original BSRI; $\underline{r}$ = .06 for Short form). For the most part, the neutral or filler items (every third item) did not load on either factor. Some notable exceptions, however, were helpful (33), sincere (57), and friendly (59), which loaded on Factor 1.

Table 7

Factor Loadings from the BSRI Original Form (60 Items) and the Short Form (30 Items)

BSRI Item	Original Form Factor 1	Original Form Factor 2	Short Form Factor 1	Short Form Factor 2
1. defend own beliefs	.15	**.36**	.13	**.39**
2. affectionate	**.61**	-.05	**.67**	-.04
3. conscientious	.41	.11	.39	.10
4. independent	.13	**.47**	.11	**.46**
5. sympathetic	**.68**	-.14	**.72**	-.09
6. moody	-.09	-.03	.01	.03
7. assertive	.10	**.66**	.11	**.71**
8. sensitive to others' needs	**.72**	-.05	**.74**	.00
9. reliable	.40	.18	.31	.16
10. strong personality	.21	**.56**	.18	**.55**
11. understanding	**.58**	-.05	**.60**	-.01
12. jealous	-.22	-.02	-.14	.03
13. forceful	-.26	**.38**	-.21	**.47**
14. compassionate	**.69**	-.05	**.75**	-.03
15. truthful	.39	.11	.31	.09
16. have leadership abilities	-.01	**.47**	-.02	**.68**
17. eager to soothe feelings	**.62**	.01	**.66**	.02
18. secretive	-.17	.07	-.13	.07
19. willing to take risks	-.01	**.47**	.01	**.42**
20. warm	**.75**	-.08	**.77**	-.08
21. adaptable	.37	.21	.36	.21
22. dominant	-.22	**.66**	-.18	**.72**
23. tender	**.70**	-.18	**.74**	-.14
24. conceited	-.26	.29	-.20	.29
25. willing to take a stand	.16	**.60**	.16	**.58**
26. love children	**.49**	-.01	**.47**	-.02
27. tactful	.35	.27	.35	.29
28. aggressive	-.19	**.68**	-.15	**.73**
29. gentle	**.73**	-.17	**.76**	-.14
30. conventional	.17	.09	.16	.09
31. self-reliant	.25	**.50**		
32. yielding	**.29**	-.03		
33. helpful	.66	.00		
34. athletic	-.12	**.25**		
35. cheerful	**.55**	.09		

Table 7 (continued)

BSRI Item	Original Form Factor 1	Original Form Factor 2	Short Form Factor 1	Short Form Factor 2
36. unsystematic	.01	.08		
37. analytical	.06	**.24**		
38. shy	**.00**	-.38		
39. inefficient	-.29	-.27		
40. makes decisions easily	.10	**.31**		
41. flatterable	**.20**	.05		
42. theatrical	.04	.24		
43. self-sufficient	.19	**.42**		
44. loyal	**.52**	.08		
45. happy	.47	.16		
46. individualistic	.15	**.39**		
47. soft-spoken	**.19**	-.43		
48. unpredictable	-.13	.17		
49. masculine	-.37	**.34**		
50. gullible	**.13**	-.22		
51. solemn	.06	.01		
52. competitive	-.16	**.46**		
53. childlike	**-.03**	-.02		
54. likable	.47	.15		
55. ambitious	.22	**.51**		
56. not use harsh language	**.30**	-.09		
57. sincere	.65	-.02		
58. act as a leader	.09	**.72**		
59. feminine	**.45**	-.28		
60. friendly	.59	.02		

Given that the above analyses demonstrated that the data from the present sample have high reliabilities and a factor structure similar to previous studies, scale scores were calculated for the BSRI Original and Short forms for all participants. The percentage of participants in each of the four classifications (feminine, masculine, androgynous, undifferentiated) was calculated and then compared to Table D-1 in the BSRI manual (Bem, 1981a), which lists Bem's corresponding data.

The median for the Femininity score (sexes combined) was 4.90 for Bem's norms and 5.05 for the present sample (Original form). For the Short form, the median for the Femininity score (sexes combined) was 5.50 for Bem's norms and 5.80 for the present sample. For the Original form Masculinity score, the median was 4.95 for Bem's normative data and 4.95 for the data derived from the present sample. The median for the Short form Masculinity score was 4.80 for Bem's norms and 4.90 for the present sample. Also, for this sample, the correlation between the Original and Short form Masculinity scores was .92; the correlation between the Original and Short form Femininity scores was .89. Table 8 presents the percentages of participants in the Bem normative sample and the present sample who were classified as "feminine," "masculine," "androgynous," and "undifferentiated" on the basis of the median split and hybrid methods. Original and Short form results are presented.

Table 8

Percentages of Participants by Category in Bem's Four-fold Classification System

	Feminine	Masculine	Androgynous	Undifferentiated
Original Form				
Median-split Method				
Females				
Bem	39.4	12.4	30.3	17.9
Hoffman	34.4	17.6	25.6	22.3
Males				
Bem	11.6	42.0	19.5	26.9
Hoffman	4.1	55.1	22.4	18.4
Hybrid Method				
Females				
Bem	41.2	10.0	24.1	24.7
Hoffman	46.2	8.8	33.3	11.7
Males				
Bem	12.2	40.8	14.1	33.0
Hoffman	10.2	45.9	26.5	17.3
Short Form				
Median-split Method				
Females				
Bem	23.8	15.6	37.1	23.5
Hoffman	32.6	19.4	27.8	19.4
Males				
Bem	16.0	32.6	23.9	27.5
Hoffman	6.1	45.9	24.5	22.4

Table 8 (continued)

	Feminine	Masculine	Androgynous	Undifferentiated
Hybrid Method				
Females				
Bem	29.4	16.2	27.4	27.1
Hoffman	39.1	11.4	44.3	5.2
Males				
Bem	14.7	28.2	19.1	38.0
Hoffman	9.3	36.1	52.6	2.1

A significant relationship was found between Bem's sample and the sample in this study when the median-split method was used (chi-square = 21.78, df = 3, p < .001). An inspection of the critical ratios indicated that there were more than the expected number of undifferentiated females and fewer than the expected number of androgynous females in the present sample. For males in this study, there were fewer than the number expected in the feminine category, and more than the number expected in the masculine category. In contrast with Bem's sample, then, more females in the present sample scored below the median on both the Feminine and the Masculine scales. More males scored above the median on the Masculine scale and below the median on the Feminine scale.

The hybrid scoring method also yielded a significant relationship between Bem's data and that of the present study (chi-square = 47.79, df = 3, p < .001).

An inspection of the critical ratios indicated that, for both males and females in the present sample, there were more than the expected number of androgynous and fewer than the expected number of undifferentiated participants when the hybrid method was used. Thus, the hybrid method yielded a significantly greater number of participants in the present study who could be seen as strong on both traditionally masculine and traditionally feminine traits, and a significantly lower number who could be labelled weak on both.

Although the numbers of participants that would be expected to comprise the various classification categories in the present sample differed significantly from the numbers that actually were in the classifications derived from this sample, the relationships between Bem's sample and the present sample were in marked contrast depending on the scoring method used. When comparing the results obtained by the hybrid method with those obtained by the median-split method, it should be noted that Bem (1981a) found that differences in classification occurred for approximately one-fourth of respondents (24% on Original form and 29% on Short form). In the present study, however, difference in scoring method resulted in a change of classification for 41% of respondents on the Original form and 39% of respondents on the Short form. Those typically affected by varying the scoring method are respondents with small Femininity-minus-Masculinity scores whose Femininity and Masculinity scores fall on opposite sides of their respective medians, as well as participants with large Femininity-minus-Masculinity scores whose Femininity and

Masculinity scores fall on the same side of their respective medians. For the present sample, those affected comprised a substantial number of participants, resulting in very different classification group sizes depending which method was used. Regardless of method, however, there was a significant difference between the percentage of college undergraduates classified as feminine, masculine, androgynous, and undifferentiated in Bem's 1978 sample and the percentage of college undergraduates classified as feminine, masculine, androgynous, and undifferentiated in this sample. Therefore, research hypothesis one was supported. Furthermore, despite Bem's (1981a) minimization of the importance of the scoring method utilized, both the median-split and the hybrid methods were used in the data analyses for this study so that possibly significant findings would not be overlooked.

Research Hypothesis Two

> There will be no significant agreement among college undergraduates supporting the "masculinity" of the items that comprise the BSRI Masculine Scale. Similarly, there will be no significant agreement among college undergraduates supporting the "femininity" of the items that comprise the BSRI Feminine scale.

Both parts of this hypothesis required the calculation of an index of neutrality level for each BSRI item. For respondents in this study, an agreement level of 75% was specified for an item to be classified as neutral, masculine, or feminine. The 75% agreement level has been used as an

indication of stereotypes in similar research (Ballard-Reisch & Elton, 1992; Broverman et al., 1972). Of the 60 BSRI items, 22 items were determined to be neutral by at least 75% of the participants. These items were: defend my own beliefs (1), conscientious (3), independent (4), reliable (9), strong personality (10), truthful (15), have leadership abilities (16), adaptable (21), willing to take a stand (25), conventional (30), self-reliant (31), helpful (33), unsystematic (36), inefficient (39), self-sufficient (43), loyal (44), happy (45), individualistic (46), solemn (51), likable (54), ambitious (55), and friendly (60). Of these 22 items, nine are from the BSRI "Masculine" scale (items 1, 4, 10, 16, 25, 31, 43, 46, and 55), one is from Bem's "Feminine" scale (item 44), and the remaining 12 items are filler or neutral items. Masculine (49) was the only one of the 60 BSRI items to reach a 75% agreement level to be classified as masculine. Similarly, feminine (59) was the only item of the 60 that qualified as feminine. The 75% agreement level was not reached for the remaining 36 BSRI items. More specific information pertaining to hypothesis two can be found in Table 9.

Because "masculine" and "feminine" were the only two of the 40 items from the BSRI Masculine and Feminine scales that were determined not to be gender-neutral, hypothesis two was supported.

Table 9

Frequency and Percentage of Respondents Evaluating BSRI Item as "Masculine," "Feminine," and "Neutral"

BSRI Item	Feminine		Masculine		Neutral	
	N	%	N	%	N	%
1. defend my own beliefs	8	2	14	4	349	94
2. affectionate	158	43	0	0	213	57
3. conscientious	69	19	16	4	284	77
4. independent	17	5	48	13	306	83
5. sympathetic	201	54	0	0	170	46
6. moody	132	36	21	6	218	59
7. assertive	12	3	132	36	227	61
8. sensitive to others' needs	207	56	0	0	164	44
9. reliable	43	12	19	5	309	83
10. strong personality	11	3	64	17	296	80
11. understanding	140	38	1	0	230	62
12. jealous	44	12	57	15	269	73
13. forceful	0	0	244	66	125	34
14. compassionate	183	50	0	0	187	51
15. truthful	53	14	1	0	317	85
16. have leadership abilities	9	2	66	18	296	80
17. eager to soothe feelings	228	62	4	1	139	38
18. secretive	82	22	51	14	237	64
19. willing to take risks	1	0	160	43	210	57
20. warm	166	45	3	1	202	54
21. adaptable	35	10	32	9	303	82
22. dominant	1	0	204	55	166	45
23. tender	209	56	1	0	161	43
24. conceited	15	4	97	26	257	70
25. willing to take a stand	9	2	61	16	301	81
26. love chidren	109	29	0	0	262	71
27. tactful	65	18	40	11	264	72
28. aggressive	1	0	193	52	177	48
29. gentle	183	49	7	2	181	49
30. conventional	34	9	30	8	304	83
31. self-reliant	8	2	47	13	314	85
32. yielding	123	33	6	2	239	65

Table 9 (continued)

BSRI Item	Feminine		Masculine		Neutral	
	N	%	N	%	N	%
33. helpful	78	21	4	1	287	78
34. athletic	2	1	98	27	269	73
35. cheerful	106	29	2	1	261	71
36. unsystematic	21	6	66	18	282	76
37. analytical	37	10	63	17	269	73
38. shy	112	30	5	1	252	68
39. inefficient	15	4	40	11	314	85
40. makes decisions easily	23	6	89	24	257	70
41. flatterable	136	37	11	3	222	60
42. theatrical	97	26	17	5	255	69
43. self-sufficient	17	5	43	12	309	84
44. loyal	80	22	13	4	276	75
45. happy	37	10	4	1	328	89
46. individualistic	14	4	36	10	319	86
47. soft-spoken	225	61	2	1	142	39
48. unpredictable	39	11	77	21	253	69
49. masculine	1	0	287	78	81	22
50. gullible	131	36	15	4	223	60
51. solemn	30	8	39	11	300	81
52. competitive	1	0	133	36	235	64
53. childlike	50	14	74	20	245	66
54. likable	31	8	4	1	334	91
55. ambitious	10	3	36	10	323	88
56. do not use harsh language	155	42	6	2	208	56
57. sincere	100	27	4	1	265	72
58. act as a leader	4	1	99	27	266	72
59. feminine	291	79	3	1	75	20
60. friendly	41	11	1	0	326	89

Research Hypothesis Three

There will be two factors that define college undergraduates' levels of gender self-confidence, identified as (a) gender self-definition and (b) gender self-acceptance.

As in the pilot study, descriptive statistics were calculated separately for males and females for each of the items on the Hoffman Gender Scale (HGS). In its revised form, the HGS is comprised of 14 items, 12 of which were included in the original form. Two new items (Items 1 and 9) were added to strengthen the anticipated self-definition factor. (See Table 1 for the original form of the HGS and Appendices A and B for the revised forms.)

Factor analyses were conducted to assess the dimensionality of gender self-confidence. Two factors, which can be identified as gender self-definition and gender self-acceptance, accounted for 62% of the variance for both the female and male respondents. As expected, HGS items 1, 4, 6, 7, 9, 12, and 14 formed the gender self-definition factor. Also as expected, the gender self-acceptance factor was defined by HGS items 2, 3, 5, 8, 10, 11, and 13. Table 10 presents the mean, standard deviation, item-total correlation (r), and the two factor loadings for each item.

Estimates of the reliability of each of the two HGS subscales (Gender Self-definition and Gender Self-acceptance) were determined separately for males and females using coefficient alpha. For females, coefficient alpha was .88 for the Self-definition subscale and .90 for the Self-acceptance subscale.

Table 10

HGS Descriptive Statistics by Sex of Respondent and Factor Loadings

	Females					Males				
Item	Mean	SD	r	F 1	F 2	Mean	SD	r	F 1	F 2
Gender Self-definition										
1. Describe	3.31	1.53	.52	**.57**	-.06	3.16	1.67	.59	**.71**	-.23
4. Biology	2.49	1.36	.56	**.50**	.15	2.21	1.34	.67	**.61**	.16
6. Define	3.43	1.42	.79	**.93**	-.08	3.15	1.56	.87	**.94**	-.03
7. Identity	3.29	1.40	.79	**.92**	-.05	2.99	1.48	.88	**.97**	-.07
9. Critical	3.02	1.36	.69	**.70**	.00	2.87	1.48	.78	**.76**	.07
12. Self-conp	2.76	1.29	.72	**.75**	.02	2.63	1.34	.81	**.81**	.13
14. Contrib	2.84	1.38	.60	**.57**	.13	2.44	1.36	.78	**.74**	.17
Gender Self-acceptance										
2. Confid	1.65	.96	.72	.09	**.68**	1.44	.69	.65	.32	**.35**
3. Stand	1.76	.98	.65	.03	**.64**	1.64	1.00	.50	.21	**.33**
5. Secure	1.64	.91	.78	.05	**.76**	1.53	.76	.63	.32	**.33**
8. Regard	1.77	.96	.63	.01	**.67**	2.00	1.07	.46	.27	**.47**
10. Happy	1.40	.77	.74	-.05	**.86**	1.43	.67	.50	-.08	**.78**
11. Comfort	1.31	.68	.69	-.02	**.78**	1.27	.51	.54	-.17	**.83**
13. Sense	1.51	.81	.78	-.07	**.89**	1.45	.59	.65	.08	**.76**

N = 371 (Females = 273, Males = 98)

F 1 = Factor 1 (Gender Self-definition)
F 2 = Factor 2 (Gender Self-acceptance)

For males, alpha was .93 for the Self-definition subscale and .80 for the Self-acceptance subscale. A MANOVA was calculated to assess a possible difference between item responses of males and females. There were no overall differences between males and females across the seven items that comprise the Gender Self-definition subscale nor across the seven items that form the Gender Self-acceptance subscale [F (2, 368) = 1.72, p = .181]. Thus, hypothesis three was supported.

Research Hypothesis Four

There will be no significant relationship between college undergraduates' levels of gender self-confidence, as measured by the Hoffman Gender Scale (HGS) and their self-descriptions according to BSRI classifications.

MANOVA was used to examine the relationship of the means of the two gender self-confidence subscales across the four classification categories used by Bem (1981a): feminine, masculine, androgynous, and undifferentiated, when male and female participants are divided into these four categories. The sample was divided, by sex, into the four classification categories by two methods. First, the median-split classification system was used based upon medians derived from this sample. The hybrid method also was used. These results are presented in Table 11. For neither classification method were there any significant differences in the mean HGS subscales related to gender, the four categories, or the interaction between gender and BSRI classification. Thus research hypothesis four was supported.

Table 11

Relationship Between HGS Scale Scores and BSRI Classification (MANOVA)

Original Form

Median-split Method

Effect	df	F	p
Gender	(2, 362)	.54	.586
Classification	(6, 724)	2.02	.060
Gender by classification	(6, 724)	.88	.513

Hybrid Method

Effect	df	F	p
Gender	(2, 362)	.31	.735
Classification	(6, 724)	.35	.912
Gender by classification	(6, 724)	2.17	.043

Short Form

Median-split Method

Effect	df	F	p
Gender	2, 359	2.56	.079
Classification	6, 718	2.12	.049
Gender by classification	6, 718	.59	.737

Hybrid Method

Effect	df	F	p
Gender	2, 359	.12	.884
Classification	6, 718	.81	.565
Gender by classification	6, 718	.95	.457

Research Hypothesis Five

There will be a significant positive relationship between college undergraduates' levels of gender self-confidence, as measured by the Hoffman Gender Scale (HGS), and their neutral evaluations of BSRI "masculine" and "feminine" items, as measured by an evaluation score established for each participant.

The evaluation score is the number of BSRI items endorsed by a given participant as "neutral." For purposes of this study, it was established as a measure of the degree of neutrality with which each respondent viewed the 60 items comprising the BSRI. For the females, the average number of items evaluated as neutral was 40 (sd = 15); for the males, the average number was 39 (sd = 16). ANOVA results indicated no difference between males and females on this measure [F (1, 369) = .87, p = .350].

For females, there was a significant correlation between their evaluation scores and their scores on the Gender Self-definition subscale of the HGS (r = .18, df = 270, p < .001). The correlation between females' evaluation scores and their scores on the HGS Gender Self-acceptance subscale, however, was not significant (r = -.04, df = 270, p > .05). For males in this study, the correlation between their evaluation scores and their HGS Gender Self-definition subscale scores also was significant (r = .28, df = 95, p < .001). As was the case for females, however, the correlation between males' evaluation scores and their HGS Gender Self-acceptance subscale scores was not

significant (r = -.02, df = 95, p > .05). Lower HGS subscale scores indicate higher levels of gender self-definition and gender self-acceptance; therefore, the positive correlations described above suggest a negative relationship between gender self-definition and perception of BSRI items as neutral.

Thus, while gender self-definition appears to be negatively related to female and male participants' neutral evaluations of the BSRI items, gender self-acceptance appears to be unrelated to such evaluations. Consequently, because neither dimension of gender self-confidence was found to be positively associated with perceived neutrality of BSRI items, hypothesis five was not supported.

CHAPTER V

SUMMARY, LIMITATIONS, RECOMMENDATIONS, IMPLICATIONS, AND CONCLUSIONS

In this final chapter, the study is summarized, conclusions are drawn, limitations are noted, recommendations are provided, and implications are discussed. These issues are addressed in the context of the literature reviewed previously and in the framework provided by the research hypotheses proposed in Chapter Three. Interpretations of the results of the data analyses are offered.

Summary

This study was a reexamination of masculinity and femininity as psychological constructs, as well as an attempt to test an alternative approach to their measurement. Five research questions and five corresponding hypotheses were developed around these components.

Research Hypothesis One

The first hypothesis considered self-descriptions of college undergraduates. By virtue of its status as the most widely used instrument of its kind, the Bem Sex-Role Inventory (BSRI; Bem, 1974) was selected as a primary focus of this study, as well as a means to examine college undergraduates' self-reported levels of characteristics that traditionally have

been associated with masculinity and femininity. Although the classifications assigned to participants in this study (i.e., masculine, feminine, androgynous, undifferentiated) varied according to which form of the BSRI (Original or Short) and which scoring method (median-split or hybrid) was used, findings indicated that current college undergraduates' self-descriptions on the BSRI were substantially different from those of college undergraduates in 1978.

As described in Chapter Four, when the median-split scoring method was used, a significantly greater number of females in this sample than in Bem's sample scored below the median on both the Masculine and the Feminine scales, thus classified as undifferentiated. More males scored above the median on the Masculine scale and below the median on the Feminine scale, thus classified as masculine. Fewer males scored above the median on the Feminine scale while scoring below the median on the Masculine scale, thus assigned to the feminine classification group. However, when the hybrid scoring method was used to classify participants, the sizes of the four groups in relation to one another were quite different. Specifically, the hybrid method resulted in a significantly greater number of androgynous males and females in this sample than in Bem's sample. Thus, for females, the hybrid method yielded opposite results from the median split method, which had resulted in more than the expected number of undifferentiated females. Therefore, although hypothesis one was supported in that current college undergraduates' self-descriptions on the BSRI were indeed different from those reported by Bem

in 1978, it remains unclear how meaningful these classifications are to begin with, in light of the inconsistencies in respondents' classifications across scoring method. Such inconsistencies have not been noted in previous research.

Research Hypothesis Two

The BSRI was further used as a vehicle by which contemporary college undergraduates' perceptions of femininity and masculinity could be assessed. Specifically, the second hypothesis tested was that current college undergraduates would not view the items that comprise the Masculine and Feminine scales of the BSRI in gender-linked terms. Consistent with the findings of Ballard-Reisch and Elton (1992), overwhelming support for this hypothesis was established, with "masculine" and "feminine" being the only items on the entire inventory which met the 75 percent agreement level necessary to be classified as such. The remaining 19 items on the BSRI Masculine scale and the remaining 19 items on the BSRI Feminine scale failed to meet this criterion.

Of the total 60 BSRI items, 22 were evaluated as neutral by at least 75 percent of the participants in this study. Among these were such traditionally masculine items as "defend my own beliefs" (94%), "ambitious" (88%), "individualistic" (86%), "self-reliant (85%), "self-sufficient" (84%), "independent" (83%), "willing to take a stand" (81%), "have leadership abilities" (80%), and "strong personality" (80%). Clearly, college undergraduates in this study perceived BSRI items very differently from the gender-stereotypical way that the

judges in Bem's test development process viewed these descriptors.

These differences give further cause to doubt the meaningfulness of the fourfold classification system by which BSRI scale scores are interpreted. If the items that comprise the BSRI Masculine scale are no longer considered masculine, and the items on the BSRI Feminine scale are no longer considered feminine, then the basis for classifying individuals in such terms is eroded. The results of the present study suggest that gender schema theory (Bem, 1981a), which relied upon cultural definitions of masculinity and femininity as a framework for one's organization of information about self and others, no longer has a foundation, if indeed it ever did.

Bem (1979) herself argued that "behavior should have no gender," and acknowledged that "the concept of androgyny contains an inner contradiction and hence the seeds of its own destruction" (p. 1053). The concept of androgyny suggested that individuals could exhibit both "masculine" and "feminine" traits. The findings described above suggest that traits are no longer perceived in those terms. Therefore, these findings suggest that, in 1996, androgyny <u>is</u> an outmoded concept, and the definitions of masculinity and femininity on which it is based are no longer relevant.

<u>Research Hypothesis Three</u>

This study was designed with the premise that masculinity and femininity could be reconceptualized in terms of gender identity (Spence, 1984, 1985). Gender identity has been described as a "secure sense or conviction of one's

own maleness or femaleness" (cf. Green, 1974; Money, 1994), thus referring to one's subjective feelings of maleness or femaleness (Basow, 1992; Golombok & Fivush, 1994). Given the emphasis in these definitions on an individual's self-concept related to gender (Spence, 1985), and given that a sense of confidence in and comfort with being a male or female appears critical to such a discussion (Lewin, 1984b), gender self-confidence was identified as a construct worth investigating toward a better understanding of masculinity and femininity.

Toward this end, the Hoffman Gender Scale (HGS; Hoffman, 1996) was developed. As described in Chapter Three, the instrument was pilot-tested and revised prior to this study. Gender self-definition and gender self-acceptance were hypothesized as two factors that define college undergraduates' levels of gender self-confidence. Results of the data analyses conducted to test this hypothesis clearly supported the existence of these two factors, which accounted for 62% of the variance for both female and male participants. Thus, hypothesis three was supported.

Research Hypothesis Four

Problems with the construct validity of the BSRI, as identified by previous researchers (e.g., Lippa, 1985; Payne, 1985; Spence, 1984, 1985, 1991) were detailed in Chapter Two. Although the BSRI has been viewed by the test developer as a measure of masculinity and femininity and assumed to be such by many researchers who use it, claims that it is essentially a measure of

instrumentality and expressiveness have been substantiated (Lippa, 1985; Spence, 1985, 1991). The present study was designed on the basis of such claims. Because the conceptualization of masculinity and femininity in terms of instrumentality and expressiveness is different from the conceptualization of masculinity and femininity as representations of gender identity, the lack of a relationship between participants' levels of gender self-confidence and their BSRI scores was hypothesized. As demonstrated in Chapter Four, this hypothesis was supported by the finding that, for neither the median-split nor the hybrid scoring method, were there any differences in the means of the HGS subscales related to gender, the four classification categories, or the interaction between gender and BSRI classification. As hypothesized, gender self-confidence, as a component of gender self-concept, and, thus, a component of gender identity, is unrelated to stereotypical descriptions of masculinity and femininity that are associated with instrumentality and expressiveness, respectively. These findings support earlier arguments that gender self-concept (Lewin, 1984b) and gender identity (Spence, 1984, 1985) are independent from conventional sex role stereotypes in defining masculinity and femininity.

Research Hypothesis Five

A significant positive relationship was hypothesized between college undergraduates' levels of gender self-confidence and their neutral evaluations of BSRI "masculine" and "feminine" items. This was the only one of the five research hypotheses not supported by the findings of the study. For both

females and males, a significant negative relationship was found between respondents' levels of gender self-definition and their tendency to perceive BSRI items as neutral. Gender self-acceptance, on the other hand, was not associated with participants' neutral evaluations of these items.

Originally, the researcher hypothesized that the more gender self-confident an individual was, the more likely that person would be to view traditionally "feminine" and "masculine" characteristics in neutral terms. It was found, however, that the more that an individual defines oneself in terms of one's masculinity or femininity (gender self-definition), the more likely that individual would be to attach "masculine" and "feminine" labels to human characteristics. Interpretation of the research findings related to hypothesis five requires further consideration of gender self-definition as the dimension of gender self-confidence that was related to neutral evaluations. The gender self-definition factor was defined by such HGS items as "My identity is strongly tied to my femininity (masculinity)," "When I am asked to describe myself, being female (male) is one of the first things I think of, and "I define myself largely in terms of my femininity (masculinity)." It may be that many of those who responded more positively to these statements adhered to more stereotypical and less personal notions of femininity and masculinity. It is quite possible that by emphasizing their "femaleness" or "maleness," they were, in fact, emphasizing those aspects of themselves that have traditionally been associated with one sex or the other.

Gender self-acceptance is defined by HGS items such as "I meet my personal standards for femininity (masculinity)," "I am secure in my femininity (masculinity)," and "My sense of myself as a female (male) is positive." These items suggest an acceptance of oneself as male or female rather than the definition of oneself as such, and, therefore, responses to these items would less likely be influenced by one's perceptions of femininity or masculinity, as found here.

Interpretations and Conclusions

Several conclusions can be drawn from the results of the study. Some are directly related to the hypotheses that formed the basis for the study, whereas others are related indirectly. All interpretations and conclusions are related to at least one of the two instruments which were central to the study, the Bem Sex-Role Inventory and the Hoffman Gender Scale.

The Bem Sex-Role Inventory

In Chapters One and Two, questions and concerns were raised regarding use of the BSRI in research. Conceptual and methodological issues were raised and discussed so that researchers might be more likely to carefully consider the implications of their use of this instrument. Here, issues related to the BSRI factor structure and the BSRI classification systems are examined further.

BSRI Factor 1. The first factor that resulted from an oblique rotation (oblimin) was defined primarily by the items that comprised Bem's (1981a)

Feminine Scale, with the exception of Original form items such as: yielding (32), shy (38), flatterable (41), soft-spoken (47), gullible (50), and childlike (53). (See Table 7 for factor loadings.) Also loading on Factor 1 were several items that Bem used as filler or neutral items. These included: helpful (33), sincere (57), and friendly (59).

Although Bem (1981a) labelled this scale "Feminine," and suggested that it was a measure of femininity, other contributors to the literature on the assessment of femininity and masculinity (e.g., Lewin, 1984b; Spence, 1984, 1985, 1991; Spence & Sawin, 1985) have concluded that the BSRI and similar instruments, such as the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974), do not measure femininity (or masculinity), and do not even measure sex-role orientation. They argued, instead, that these tests merely assess personality attributes that are conceptually independent of gender. Furthermore, such scholars (e.g., Lewin, 1984b; Spence, 1985, 1991) have recommended that the labels, Feminine (and Masculine), be rejected and replaced by terms that describe the actual content of these scales. The content of the BSRI Feminine scale has been described as representing expressiveness (Lippa, 1985; Pedhazur & Tetenbaum, 1979; Spence, 1985, 1991), warmth, expressiveness, interpersonal orientation, and sensitivity to others (Lott, 1990), and expressive(ness) in large part [as a] euphemism for female subordination (Lewin, 1984b). Ironically, perhaps the most convincing argument for not labelling the BSRI Feminine scale as "feminine" was provided

by Bem (1985) herself when she contended that "human behaviors and personality attributes should no longer be linked with gender" (p. 222).

The traits that appear on the BSRI Feminine scale are, and for over a decade, have been, conceptualized as descriptive of expressiveness by many of the leading researchers in the field. That they are not conceptualized as "feminine" was strongly supported by the findings related to research hypothesis two in the present study. College undergraduates did not view the items on the BSRI Feminine scale as feminine, with the exception of the item called "feminine" (59). Thus, although the BSRI Feminine scale has instilled confidence as a reliable measure (in this study, alpha = .84), it can be concluded that it is, at best, a reliable measure of expressiveness, not femininity.

BSRI Factor 2. An even greater number of Bem's Masculine scale items loaded successfully on Factor 2 than the number of her Feminine scale items that loaded on Factor 1. As presented in Chapter Four, the most notable exceptions were athletic (34) and analytical (37). Furthermore, Bem's filler or neutral items did not load on Factor 2 as some had on Factor 1. Factor loadings are provided in Table 7.

The BSRI Masculine scale has been subject to the same criticism as that directed at the Feminine scale. Whereas Bem (1981a) developed it as a measure of masculinity, various other researchers have viewed it differently. The personality attributes described by the items on the BSRI Masculine scale

are primarily related to instrumentality (Lippa, 1985; Pedhazur & Tetenbaum, 1979; Spence, 1985, 1991). As described in Chapter Two, some of the factor analyses performed on the BSRI have resulted in more than one instrumentality factor. Lippa's (1985) review suggested that two highly correlated instrumentality factors exist, one which can be labelled "dominance" and the other "self-reliance." Similarly, Pedhazur and Tetenbaum (1979) identified two highly correlated factors related to instrumentality, which they called "assertiveness" and "self-sufficiency." The present study supported a two-factor structure for the BSRI as a whole, one defined largely by Bem's "masculine" items and one defined largely by her "feminine" items; however, the question of what is being measured by the Masculine, as well as the Feminine scale, must be addressed. Based upon the studies described above, it would seem that the BSRI Masculine scale is a measure of instrumentality.

The characteristics that appear as items on the BSRI Masculine scale were not perceived as "masculine" by the participants in the present study, with the exception of the item called "masculine" (49). In fact, nine of the 20 items that comprise the BSRI Masculine scale met the 75% agreement level to be classified as neutral items. Thus, despite the impressive reliability indicated in the present study by an alpha coefficient of .84, the BSRI Masculine scale is more accurately viewed as a measure of instrumentality, not masculinity.

Bem herself (1978) used the words "expressive" and "instrumental" to describe the dimensions described by her Feminine and Masculine scales,

respectively. She stated that "the feminine male is low in the instrumental domain, and the masculine male is low in the expressive domain," and that "the masculine woman is low in the expressive domain, and the feminine woman is low in the instrumental domain" (p. 18). The leap from "expressiveness" to "femininity," and from "instrumentality" to "masculinity" appears unfounded.

Median-split classification system. As discussed in Chapter Four, relative BSRI classifications of the college undergraduates in this study differed greatly from those of the college undergraduates on which the BSRI was normed. The relevance of this finding, however, is in question. The medians of the Masculinity and Femininity scores were not that different between the two samples. Recall that the median for the Masculinity score (sexes combined) was 4.95 for Bem's normative data and 4.95 for the data derived from this sample. The median for the Short form Masculinity score was 4.80 for Bem's norms and 4.90 for the present sample. The median for the Femininity score (sexes combined) was 4.90 for Bem's norms and 5.05 for the present sample (Original form). For the Short form, the median for the Femininity score was 5.50 for Bem's norms and 5.80 for the present sample. These data would suggest that, in general, participants in the current sample rated themselves stronger on traditionally feminine characteristics than participants in Bem's sample, and equal to or slightly stronger on traditionally masculine characteristics. However, when classified as feminine, masculine, androgynous, or undifferentiated by the medians of the present sample, female

participants in this study were overrepresented in the undifferentiated group and underrepresented in the androgynous group. Male participants were overrepresented in the masculine category and underrepresented in the feminine category. These inconsistencies between the current sample's participants' relative strengths of "feminine" and "masculine" traits and their actual classifications cause some concern.

As indicated in Chapter Two, Spence and Helmreich (1978) contended that the median-split technique results in data subject to statistical distortion. Spence and Helmreich argued that, particularly when research questions involve between-group comparisons, such results must be viewed with considerable caution. Although Bem (1981a) acknowledged that "problematic cases" could result from the median-split method, she stated that they are "all...individuals who score near the cutoff point for femininity or masculinity or both" and merely "constitute an additional source of 'noise' or 'error' in any research design" (p. 9). The results of this study suggest that there are indeed problems inherent in the median-split classification system that require greater attention than given to this scoring method by Bem. As might be expected, the Masculinity and Femininity scale scores of a considerable number of participants in this study were close enough to the median to affect classification. This, in combination with the observations that researchers who use the BSRI seem to consistently attach a considerable degree of importance to the BSRI classification of their respondents, as do many respondents

themselves, suggests that Bem's perspective here is a serious minimization.

Median-split versus hybrid method. To add to this inconsistency, very different classifications emerge when the hybrid method, as opposed to the median-split method, was used to classify participants in this study. As discussed in the previous chapter, method affected classification for 41% of participants on the Original form and 39% of participants on the Short form. For both males and females in the present sample, the hybrid method yielded a significantly greater number of androgynous participants and a significantly lower number of undifferentiated participants than that which would be expected compared to Bem's sample.

Although the hybrid method resulted in groups whose relative sizes were more consistent with the norms of this study, this method is not recommended by Bem (1981a) because it is difficult to "execute and to explain" (p. 65). Bem (1981a) argued that "[a]t the present time, it is not known if one of the classification methods has greater predictive utility than the other" and that "[b]oth appear to be perfectly adequate for research" (p. 65). In the current study, classification of participants was largely affected by the method used. In fact, 20% of the participants had three different classifications depending on whether the Original or Short BSRI was scored by the median-split or hybrid method. One respondent was described by all four of the four possible classifications, depending on the form and the method. That respondent, a male, was classified as feminine on the Original form of the BSRI by the

median-split method, undifferentiated on the Original form by the hybrid method, masculine on the Short form by the median-split method, and androgynous on the Short form by the hybrid method. At the very least, then, results of this study suggest that data obtained from the use of the BSRI need to be viewed cautiously.

The Hoffman Gender Scale

Results of the study supported the use of the HGS as a tool to assess gender self-confidence. The HGS is not intended as a measure of global masculinity and femininity. Rather, it was designed to measure gender self-confidence as one component of gender self-concept, which in turn, is but one aspect of gender identity. As discussed throughout this dissertation, the gender identity construct provides a way to reconceptualize the constructs of masculinity and femininity, and to allow for an individual's personal definitions of these terms as opposed to assuming acceptance of their traditional, stereotypical meanings.

Gender self-definition and gender self-acceptance were identified as two aspects of gender self-confidence that may be assessed individually. Each of these constructs must be considered independently in order to understand what is being measured by this instrument.

Gender self-definition. As the items that comprise the Gender Self-definition subscale indicate (see Appendices A and B, items 1, 4, 6, 7, 9, 12, and 14), gender self-definition relates to how strong a component of one's

identity one considers one's femininity or masculinity to be. As such, how one defines femininity or masculinity is left up to the individual. Spence (1985) suggested that individuals create their own standards or "calculus" for self-assessing femaleness or maleness. For example, women whose career aspirations were traditionally "masculine" did not consider themselves to be "masculine" (Tangri, 1972). They defined their femininity in a variety of other ways. Similarly, women who rate themselves as strong on many of the "masculine" BSRI items, and who may be classified as masculine or androgynous by BSRI standards, may define themselves as quite feminine, and consider femininity to be an integral aspect of their self-concepts. Other women who exhibit many of those same traits may not consider femininity nearly as salient to their definitions of self. The same, of course, can apply to men, in that definitions of masculinity vary among individuals, as does the importance that masculinity has in one's definition of self as a male.

It is interesting that level of gender self-definition was found to be negatively related to participants' evaluations of BSRI items as neutral, a finding opposite to that posited by hypothesis five. This finding was interpreted to indicate that participants' definitions of masculinity and femininity to which they were referring when completing the HGS were traditional or stereotypical definitions. A related point of view which can be considered here is that definitions of masculinity and femininity, stereotypical or personal, can represent a limited, and possibly, even an unhealthy, way to define oneself. As early as

1975, Bem (1978) proposed the following "prescription" for a "liberated" identity: "Let gender move from figure to ground" (p. 21). Bem (1993) contended that, for her, being female, like being human, was a fact, a "taken-for-granted background fact rather than a nucleus around which I have constructed my identity" (p. viii). Furthermore, Bem (1993) argued that it was virtually impossible to conceive of male and female, masculinity and femininity, as notions that are independent of the hidden assumptions that she referred to as "the lenses of gender" (p. 2). She suggested that "the lenses of gender are embedded in cultural discourses, social institutions, and individual psyches in virtually all male-dominated societies" (p. 3). The point here is that, if this is so, it may be presumptuous to believe that we can even have a definition of femininity or masculinity that is separate from society, much less form a healthy identity around that definition. This argument, however, appears inconsistent with Bem's (1993) own statement that it is her "subjective sense of being outside the categories of my culture that has most profoundly contributed to my feminist politics" (p. viii). I would argue that it is precisely this sort of subjectivity that gender self-definition can be about.

Gender self-acceptance. Gender self-acceptance is related to how comfortable an individual is as a member of his or her gender. (See Appendices A and B for the particular items of this HGS subscale, items 2, 3, 5, 8, 10, 11, and 13.) Perhaps one way to conceptualize the difference between gender self-definition and gender self-acceptance is in terms of intensity.

Individuals who score low on the Gender Self-definition subscale (recall that lower scores indicate stronger gender self-definition) attribute a great deal of importance to femininity or masculinity as a part of their identity as females or males, respectively. Individuals who score low on the Gender Self-acceptance scale (recall that lower scores indicate stronger gender self-acceptance) may be able to be more relaxed about themselves as males or females, accepting themselves as such without necessarily strongly defining themselves in terms of masculinity and femininity. Thus, one might be comfortable with one's gender (gender self-acceptance), one might define oneself in terms of one's gender (gender self-definition), both, or neither.

Gender self-acceptance seems to be what Bem was describing in her 1975 address as the keynote speaker at a conference entitled "New Directions for Research on Women," planned and facilitated by the Task Force for a Conference on Women's Research Needs in Psychology of the American Psychological Association Committee on Women (Sherman & Denmark, 1978). Bem (in Sherman & Denmark, 1978) looked toward the day when "a healthy regard and acceptance of one's maleness or femaleness" would be the focus for an individual, rather than "traditional sex roles that restrict behavior" (p. xvi). In her closing remarks, Bem (1978) contended that:

> ...a healthy sense of one's maleness or femaleness becomes all the more possible precisely when the artificial constraints of gender are eliminated and one is free to be one's own unique blend of temperament and behavior. When gender no longer functions as a prison, then and

> only then will we be able to accept as given the fact that we are male or female in exactly the same sense that we accept as given the fact that we are human. (p. 21)

Gender self-confidence. Gender self-confidence, including self-definition and gender self-acceptance, is independent of sexual orientation. As Spence (1985) suggested, many lesbians and gay men define their femininity and masculinity totally separate from sexual orientation and feel confident as females and males, despite the emphasis that many heterosexual individuals place on sexual orientation in evaluating their own and others' femininity and masculinity. The importance of one's masculinity or femininity to one's self-concept can be strong, moderate, or weak regardless of one's sexual orientation. For lesbians and gay men, as well as heterosexual men and women, definitions of femininity and masculinity may vary widely. Furthermore, one may or may not accept oneself and be comfortable with one's femininity or masculinity as a lesbian, a gay male, or a heterosexual individual.

Lewin (1984b) suggested that masculinity and femininity should be assessed by measuring individuals' gender self-confidence. She argued that one's beliefs about whether one is "living up to" various aspects of one's gender-related self-concept must be addressed (p. 200). Lewin contended that confidence that one is meeting one's own standards of masculinity or femininity and that one is competent as a member of one's own sex are what need to be considered. The Hoffman Gender Scale was developed from these

contentions. It does seem that the content of what Lewin was describing is addressed more specifically by items that appear on the HGS Self-acceptance scale, rather than on the HGS Self-definition scale. However, both dimensions appear to be viable aspects of gender self-confidence that merit further investigation.

As stated earlier, gender self-definition and gender self-acceptance are two constructs which appear to define gender self-confidence. Gender self-confidence is one aspect of gender self-concept. Gender self-concept is one component of gender identity. Gender identity is a key to untangling the concepts of masculinity and femininity. This way of conceptualizing masculinity and femininity is consistent with that of Ashmore (1990), who acknowledged the complexity of the femininity and masculinity constructs and supported a multifaceted approach to understanding gender. Thus, gender self-confidence is considered here as only one step, albeit an important one, toward that understanding. Definitions of masculinity and femininity require additional attention, however, which may lead to a more widely accepted view of their complexity (Burnett, Anderson, & Heppner, 1995). Such attention may be the focus of additional research.

Limitations of the Study and Recommendations for Future Research

Results of this study need to be viewed with certain limitations in mind. Some of these limitations form the basis for future studies. Others are less amenable to being addressed by subsequent research. In this section,

limitations are noted, followed by recommendations for future research that stem from some of the limitations as well those that emerge from the findings themselves.

Limitations and Related Recommendations for Future Research

Primarily, limitations concern the generalizability of the study's results. Because participation was voluntary, it remains unknown how responses of those who did not participate might have differed from those who did. This limitation is inevitable based upon participants' right to choose whether they wish to be involved in a particular study. Generalizability is further limited by the types of classes to which the assessments were administered, and is restricted to the geographical region in which the study was conducted. This could be addressed by similar studies conducted with members of classes in other departments and schools within other academic settings.

An additional limitation may stem from reliance on self-report measures. Respondents were asked to rate themselves in two of the three aspects of the study, while the third was based on subjective ratings of human characteristics. It should be noted, however, that self-report measures can sometimes provide more dependable estimates of personality-related variables than can behavioral measures (Hattie, 1992; Howard, 1990; Howard, Maxwell, Weiner, Boynton, & Rooney, 1980). Furthermore, Lewin (1984b) and Spence (1985), from whose theoretical perspectives much of this study was developed, were adamant that individuals' own beliefs and sense of self as female or male are the issue, not

what someone else thinks of them. Thus, this particular limitation also may be viewed as a strength. Depending on one's point of view, then, this limitation may or may not require acknowledgement.

Another possible limitation may be some effect of respondents being required to read through the items on the BSRI twice, initially as a self-report, followed by assessing each item for its femininity, masculinity, or neutrality. Although varying the order of the instruments was an attempt to mitigate this type of effect, it is possible that some participants may have attended less conscientiously to the BSRI items the second time they were presented with them. Subsequent studies may be conducted in which the evaluation component is examined without participants completing the BSRI as a self-description, particularly in light of the questionable meanings that can be attributed to BSRI results.

Recommendations for Future Research Based on Findings of the Study

The focus of this study has been on the individual's gender self-confidence as a personal aspect of one's gender self-concept. Additional research can be conducted to facilitate identification and understanding of other emotional and physical components of gender self-concept. These might include constructs such as presentation of self, found in Song and Hattie's model of general self-concept (Hattie, 1992).

Furthermore, because one's self-concept is not developed or maintained in a vacuum, it would be important to consider the social components of one's

gender self-concept, specifically, how one's gender self-concept is related to interactions with one's peers, one's family, one's co-workers, and other significant people in one's life. A basic tenet of Bem's theory was that the traits that individuals display in various situations differ according to the situation; the behaviors in which one engages are often specific to the situation in which one finds oneself. Perhaps one's gender self-concept, like self-concept in general, is largely a function of the particular social setting. If this is true, a contextual approach to measurement of gender self-concept would be necessary to supplement the individualistic perspective that gender self-confidence implies.

As indicated in Chapter Three, the HGS concluded with the question, "What do *you* mean by femininity (masculinity)?" Individual responses to this question can be linked with HGS subscale scores to provide additional information about how various people define and accept themselves related to their gender. This type of follow-up study is particularly relevant in light of the findings pertaining to research hypothesis five.

Finally, subsequent studies can be developed to address the research question "How does gender self-confidence affect behavior?" Learning more about the relationship between an individual's level of gender self-confidence and his or her actions, as well as the relationship between other aspects of one's gender self-concept and gender identity and one's actions, may have implications for counseling practice.

Implications for Counseling Practice and Counselor Education

The very labelling of certain qualities as feminine or masculine encourages people to view human characteristics dichotomously, which can lead to selective perception and distortions of the actual behaviors of women and men (Enns, 1994). Like everyone else, counselors are not immune to these pitfalls. This study indicates that views of what is "masculine" and what is "feminine" have changed, lending additional support to counselors' being more intentional in their reinforcement of non-traditional and non-stereotypical gender perspectives. It is important that counselors "convey to clients the complexity and diversity of normal human behavior and encourage them to think creatively about how they want to define themselves" (Enns, 1994, p. 131). This message can be conveyed in a variety of ways, such as role-modeling, bibliotherapy, and group counseling, to name a few.

Perhaps the most salient implication concerns practitioners' and educators' own levels of self-awareness of the gender-related messages they convey. There is potential for harm to clients and students when counselors and counselor educators are concerned but unaware of the extent of their own gender biases, stereotypes, and issues that get played out in the professional arena. Moreover, if counselors, counselor educators, and counseling supervisors (Rigazio-Digilio, Anderson, & Kunkler, 1995) are not able to relinquish the expert role and model an openness to self-examination of gender attitudes, chances are likely that their clients, students, and supervisees will

experience similar difficulty. As suggested in Chapter One, this study was intended as a necessary step toward awareness and elimination of gender-related restrictions which derive from stereotypical attitudes and behavior. The outcomes of this study support the challenge to counselors and counselor educators to reexamine their gender attitudes, and to consider whether their expectations and behaviors are differentially attributed to females and males based on outmoded cultural norms.

Conclusion

Has androgyny become the outmoded concept that Bem predicted and hoped it would become? The findings of this study suggest that it may have. In 1975, Bem (1978) spoke of the day "when androgyny becomes reality, [when] the concept of androgyny will have been transcended" (p. 19). Bem's research (e.g., Bem, 1974, 1975, 1981a, 1985; Bem & Lenney, 1976; Bem, Martyna, & Watson, 1976) demonstrated that traditional gender roles restricted behavior of males and females and that it was inhibiting for individuals to adhere to what has been considered "appropriate" gender-role behavior (Sherman & Denmark, 1978). In 1996, traditional conceptions of masculinity and femininity appear to be defunct. The disregard, if not transcendence, of gender roles is becoming increasingly evident. The demise of gender roles can allow for investigations of more critical aspects of our gendered selves. Gender self-acceptance, and perhaps to a somewhat lesser degree, gender self-definition, have potential as two such aspects.

REFERENCES

Antill, J. K., Cunningham, J. D., Russell, G., & Thompson, N. L. (1981). An Australian sex-role scale. Australian Journal of Psychology, 33, 169-183.

Antill, J. K., & Russell, G. (1982). The factor structure of the Bem Sex-Role Inventory: Method and sample comparisons. Australian Journal of Psychology, 34, 183-193.

Ashmore, R. D. (1990). Sex, gender, and the individual. In L. A. Pervin (Ed.), Handbook of personality theory and research (pp. 486-526). New York: Guilford.

Ballard-Reisch, D., & Elton, M. (1992). Gender orientation and the Bem Sex Role Inventory: A psychological construct revisited. Sex Roles, 27, 291-306.

Basow, S. A. (1992). Gender: Stereotypes and roles (3rd ed.). Monterey, CA: Brooks/Cole.

Baldwin, A. C., Stevens, L. C., Critelli, J. W., & Russell, S. (1986). Androgyny and sex role measurement: A personal construct approach. Journal of Personality and Social Psychology, 51, 1081-1088.

Beere, C. A. (1979). Women and women's issues: A handbook of tests and measures. San Fransisco: Jossey-Bass.

Beere, C. A. (1990). Gender roles: A handbook of tests and measures. New York: Greenwood.

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

Bem, S. L. (1975). Sex role adaptability: One consequence of psychological androgyny. Journal of Personality and Social Psychology, 31, 634-643.

Bem, S. L. (1977). On the utility of alternative procedures for assessing psychological androgyny. Journal of Clinical and Consulting Psychology, 45, 196-205.

Bem, S. L. (1978). Beyond androgyny: Some presumptuous prescriptions for a liberated sexual identity. In J. A. Sherman & F. L. Denmark (Eds.), The psychology of women: Future directions in research (pp. 1-23). New York: Psychological Dimensions.

Bem, S. L. (1979). Theory and measurement of androgyny: A reply to the Pedhazur-Tetenbaum and Locksley-Colten critiques. Journal of Personality and Social Psychology, 37, 1047-1054.

Bem, S. L. (1981a). Bem Sex-Role Inventory: Professional manual. Palo Alto, CA: Consulting Psychologists Press.

Bem, S. L. (1981b). The BSRI and gender schema theory: A reply to Spence and Helmreich. Psychological Review, 88, 354-364.

Bem, S. L. (1981c). 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 (pp. 179-226). Lincoln, NB: University of Nebraska Press.

Bem, S. L. (1987). Gender schema theory and the romantic tradition. In P. Shaver & C. Hendrick (Eds.), Review of personality and social psychology (Vol. 7, pp. 251-271). Newbury Park, CA: Sage.

Bem, S. L. (1993). The lenses of gender: Transforming the debate on sexual inequality. New Haven, CT: Yale University Press.

Bem, S. L., & Lenney, E. (1976). Sex typing and the avoidance of cross-sex behavior. Journal of Personality and Social Psychology, 33, 48-54.

Bem, S. L., Martyna, W., & Watson, C. (1976). Sex typing and androgyny: Further explorations of the expressive domain. Journal of Personality and Social Psychology, 34, 1016-1023.

Block, J. H. (1973). Conceptions of sex role: Some cross-cultural and longitudinal perspectives. American Psychologist, 28, 512-526.

Boldizar, J. P. (1991). Assessing sex typing and androgyny in children: The Children's Sex Role Inventory. Developmental Psychology, 27, 505-515.

Brannon, R. (1978). Measuring attitudes toward women (and otherwise): A methodological critique. In J. A. Sherman & F. L. Denmark (Eds.), The psychology of women: Future directions in research (pp. 645-731).

New York: Psychological Dimensions.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex role stereotypes: A current appraisal. Journal of Social Issues, 28, 59-78.

Burnett, J. W., Anderson, W. P., & Heppner, P. P. (1995). Gender roles and self-esteem: A consideration of environmental factors. Journal of Counseling and Development, 73, 323-326.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring. Minneapolis, MN: University of Minneapolis Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

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

Cook, E. P. (1985). Psychological androgyny. New York: Pergamon.

Cook, E. P. (1987). Psychological androgyny: A review of the research. The Counseling Psychologist, 15, 471-513.

Cook, E. P. (Ed.). (1993). Women, relationships, and power:

Implications for counseling. Alexandria, VA: American Counseling Association.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Orlando, FL: Harcourt Brace Jovanovich.

Cronbach, L. J. (1960). Essentials of psychological testing (2nd ed.). New York: Harper.

Deaux, K. (1987). Psychological constructions of masculinity and femininity. In J. M. Reinisch, L. A. Rosenblum, & S. A. Sanders (Eds.), Masculinity/Femininity: Basic perspectives (pp. 289-303). New York: Oxford University Press.

Edwards, A. L., & Ashworth, C. D. (1977). A replication study of item selection for the Bem Sex Role Inventory. Applied Psychological Measurement, 1, 501-507.

Enns, C. Z. (1994). Archetypes and gender: Goddesses, warriors, and psychological health. Journal of Counseling and Development, 73, 127-133.

Frable, D. E. S., & Bem, S. L. (1985). If you're gender-schematic, all members of the opposite sex look alike. Journal of Personality and Social Psychology, 49, 459-468.

Franck, K., & Rosen, E. (1949). A projective test of masculinity-femininity. Journal of Consulting Psychology, 13, 247-256.

Garnets, L., & Pleck, J. H. (1979). Sex role identity, androgyny, and sex role transcendence: A sex role strain analysis. Psychology of Women Quarterly, 3, 270-283.

Gaudreau, P. (1977). Factor analysis of the Bem Sex-Role Inventory. Journal of Consulting and Clinical Psychology, 45, 299-302.

Gilbert, L. A. (1985). Measures of psychological masculinity and femininity: A comment on Gaddy, Glass, and Arnkoff. Journal of Counseling Psychology, 32, 163-166.

Golombok, S., & Fivush, R. (1994). Gender development. New York: Cambridge University Press.

Good, G. E., Gilbert, L. A., & Scher, M. (1990). Gender aware therapy: A synthesis of feminist therapy and knowledge about gender. Journal of Counseling and Development, 68, 376-380.

Gough, H. G. (1952). Identifying psychological femininity. Educational and Psychological Measurement, 12, 427-439.

Gough, H. G. (1964). California Psychological Inventory: Manual. Palo Alto, CA: Consulting Psychologists Press.

Green, R. (1974). Sexual identity conflict in children and adults. New York: Basic Books.

Guilford, J. P., & Martin, H. G. (1943). An Inventory of Factors GAMIN. Beverly Hills, CA: Sheridan Supply.

Guilford, J. P., & Zimmerman, W. S. (1949). The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations. Beverly Hills, CA: Sheridan Supply.

Hare-Mustin, R. T., & Maracek, J. (1994). Asking the right questions:

Feminist psychology and sex differences. Feminism and Psychology, 4, 531-537.

Harris, A. C. (1994). Ethnicity as a determinant of sex role identity: A replication study of item selection for the Bem Sex Role Inventory. Sex Roles, 31, 241-273.

Hathaway, S. R., & McKinley, J. C. (1943). The Minnesota Multiphasic Personality Inventory. New York: Psychological Corporation.

Hattie, J. (1992). Self-concept. Hillsdale, NJ: Lawrence Erlbaum.

Heerboth, J. R., & Ramanaiah, N. V. (1985). Evaluation of the BSRI masculine and feminine items using desirability and stereotype ratings. Journal of Personality Assessment, 49, 264-270.

Hefner, R., Rebecca, M., & Oleshansky, B. (1975). Development of sex role transcendence. Human Development, 18, 143-158.

Hoffman, R. M. (1996). Gender self-confidence: An innovative approach to the measurement of masculinity and femininity. Manuscript in preparation.

Howard, G. S. (1990). On the construct validity of self-reports: What do the data say? American Psychologist, 45, 292-294.

Howard, G. S., Maxwell, S. E., Weiner, R. L., Boynton, K. S., & Rooney, W. M. (1980). Is a behavioral measure the best estimate of behavioral parameters? Perhaps not. Applied Psychological Measurement, 4, 293-311.

Kelly, G. A. (1955). The psychology of personal constructs. New York:

Norton.

Kimmel, M. S., & Messner, M. A. (Eds.). (1995). Men's lives (3rd. ed.). Needham Heights, MA: Allyn & Bacon.

Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. Journal of Educational Measurement, 25, 85-96.

Kurdek, L. A., & Siesky, A. E. (1980). Sex role self-concepts of single divorced parents and their children. Journal of Divorce, 3, 249-261.

Lam, T. C. M., & Stevens, J. J. (1994). Effects of content polarization, item wording, and rating scale width on rating response. Applied Measurement in Education, 7, 141-158.

Lenney, E. (1991). Sex roles: The measurement of masculinity, femininity, and androgyny. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (Vol. 1, pp. 573-660). San Diego, CA: Academic Press.

Lewin, M. (Ed.). (1984a). "Rather worse than folly?": Psychology measures femininity and masculinity, 1. In M. Lewin (Ed.), In the shadow of the past: Psychology portrays the sexes (pp. 155-178). New York: Columbia University Press.

Lewin, M. (Ed.). (1984b). Psychology measures femininity and masculinity, 2: From "13 gay men" to the instrumental-expressive distinction. In M. Lewin (Ed.), In the shadow of the past: Psychology portrays the sexes (pp. 179-204). New York: Columbia University Press.

Lewin, M., & Wild, C. L. (1991). The impact of the feminist critique on tests, assessment, and methodology. Psychology of Women Quarterly, 15, 581-596.

Lippa, L. (1985). Review of Bem Sex-Role Inventory. In J. V. Mitchell (Ed.), The ninth mental measurements yearbook (Vol. 1, pp. 176-178). Lincoln, NB: Buros Institute of Mental Measurements.

Locksley, A., & Colten, M. E. (1979). Psychological androgyny: A case of mistaken identity? Journal of Personality and Social Psychology, 37, 1017-1031.

Lott, B. (1981). A feminist critique of androgyny: Toward the elimination of gender attributions for learned behavior. In C. Mayo & N. M. Henley (Eds.), Gender and nonverbal behavior (pp. 171-180). New York: Springer.

Lott, B. (1990). Dual natures or learned behavior: The challenge to feminist psychology. In R. T. Hare-Musten & J. Maracek (Eds.), Making a difference: Psychology and the construction of gender (pp. 65-101). New Haven , CT: Yale University Press.

Markus, H., Crane, M., Bernstein, S., & Siladi, M. (1982). Self-schemas and gender. Journal of Personality and Social Psychology, 42, 38-50.

Marsh, H. W., & Myers, M. (1986). Masculinity, femininity, and androgyny: A methodological and theoretical critique. Sex Roles, 14, 397-430.

Martin, H. J., & Ramanaiah, N. V. (1988). Confirmatory factor analysis

of the Bem Sex-Role Inventory. Psychological Reports, 62, 343-350.

McCreary, D. R. (1990). Multidimensionality and the measurement of gender role attributes: A comment on Archer. British Journal of Social Psychology, 29, 265-272.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Eductional measurement (3rd ed.) (pp. 13-103). New York: American Council on Education/Macmillan.

Mintz, L. B., & O'Neil, J. M. (1990). Gender roles, sex, and the process of psychotherapy: Many questions and few answers. Journal of Counseling and Development, 68, 381-387.

Money, J. (1994). The concept of gender identity disorder in childhood and adolescence after 39 years. Journal of Sex and Marital Therapy, 20, 163-177.

Morawski, J. G. (1985). The measurement of masculinity and femininity: Engendering categorical realities. Journal of Personality, 53, 196-223.

Morawski, J. G. (1987). The troubled quest for masculinity, femininity, and androgyny. In P. Shaver and C. Hendrick (Eds.), Review of personality and social psychology (Vol. 7, pp. 44-69). Newbury Park, CA: Sage.

Myers, A. M., & Gonda, G. (1982). Empirical validation of the Bem Sex-Role Inventory. Journal of Personality and Social Psychology, 43, 304-318.

Payne, F. D. (1985). Review of Bem Sex-Role Inventory. In J. V. Mitchell (Ed.), The ninth mental measurements yearbook (Vol. 1, pp. 178-179).

Lincoln, NB: Buros Institute of Mental Measurements.

Pedazur, E. J., & Tetenbaum, T. J. (1979). Bem Sex Role Inventory: A theoretical and methodological critique. Journal of Personality and Social Psychology, 37, 996-1016.

Pleck, J. (1984). The theory of male sex role identity: Its rise and fall, 1936 to the present. In M. Lewin (Ed.), In the shadow of the past: Psychology portrays the sexes (pp. 205-225). New York: Columbia University Press.

Rabinowitz, F. E., & Cochran, S. V. (1994). Man alive: A primer of men's issues. Pacific Grove, CA: Brooks/Cole.

Ramanaiah, N. V., & Hoffman, S. C. (1984). Effects of instructions and rating scales on item selection for the BSRI scales. Journal of Personality Assessment, 48, 145-152.

Ravinder, S. (1987). Androgyny: Is it really the product of educated, middle-class Western societies? Journal of Cross-Cultural Psychology, 18, 208-220.

Rebecca, M., & Hefner, R. (1979). The future of sex roles. In M. Richmond-Abbott (Ed.), The American woman: Her past, her present, her future (pp. 243-264).

Rebecca, M., Hefner, R., & Oleshansky, B. (1976). A model of sex-role transcendence. Journal of Social Issues, 32, 197-206.

Rigazio-DiGilio, S. A., Anderson, S. A., & Kunkler, K. P. (1995). Gender-aware supervision in marriage and family counseling and therapy: How

far have we actually come? Counselor Education and Supervision, 34, 344-355.

Sherif, C. (1982). Needed concepts in the study of gender identity. Psychology of Women Quarterly, 6, 375-398.

Sherman, J. A., & Denmark, F. L. (1978). Overview. In J. A. Sherman & F. L. Denmark (Eds.), The psychology of women: Future directions in research (pp. xiii-xxii). New York: Psychological Dimensions.

Spence, J. T. (1984). Masculinity, femininity, and gender-related traits: A conceptual analysis and critique of current research. In B. A. Maher & W. B. Maher (Eds.), Progress in experimental research in personality (Vol. 13, pp. 1-97). New York: Academic Press.

Spence, J. T. (1985). Gender identity and its implications for the concepts of masculinity and femininity. In T. B. Sonderegger (Ed.), Nebraska symposium on motivation: Psychology of gender (pp. 59-96). Lincoln, NB: University of Nebraska Press.

Spence, J. T. (1991). Do the BSRI and PAQ measure the same or different concepts? Psychology of Women Quarterly, 15, 141-165.

Spence, J. T. (1993). Gender-related traits and gender ideology: Evidence for a multifactorial theory. Journal of Personality and Social Psychology, 64, 624-635.

Spence, J. T., Deaux, K., & Helmreich, R. L. (1985). Sex roles in contemporary American society. In G. Lindzey & E. Aronson (Eds.), Handbook

of social psychology (3rd ed.) (Vol. 2, pp. 149-178). New York: Random House.

Spence, J. T., & Helmreich. R. L. (1972). The Attitudes toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. Catalog of Selected Documents in Psychology, 2, 66-67, MS 153.

Spence. J. T., & Helmreich, R. L. (1978). Masculinity & femininity: Their psychological dimensions, correlates, & antecedents. Austin, TX: University of Texas Press.

Spence, J. T., & Helmreich, R. L. (1981). Androgyny versus gender schema: A comment on Bem's gender schema theory. Psychological Review, 88, 365-368.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity-femininity. JSAS Catalog of Selected Documents in Psychology, 4, 43-44.

Spence, J. T., Helmreich, R. L., & 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 and Social Psychology, 32, 29-39.

Spence, J. T., & Sawin, L. L. (1985). Images of masculinity and femininity: A reconceptualization. In V. E. O'Leary, R. K. Unger, & B. S. Wallston (Eds.), Women, gender, and social psychology (pp. 35-66). Hillsdale,

NJ: Lawrence Erlbaum.

Strong, E. K. (1927). Strong Vocational Interest Blank. Palo Alto, CA: Consulting Psychologists Press.

Tangri, S. S. (1972). Determinants of occupational role innovation among college women. Journal of Social Issues, 28, 177-199.

Terman, L. M., & Miles, C. C. (1936). Sex and personality. New York: McGraw-Hill.

Thomas, S., & Robinson, M. (1981). Development of a measure of androgyny for young adolescents. Journal of Early Adolescence, 1, 195-209.

Thorndike, R. L., & Hagen, E. P. (1977). Measurement and evaluation in psychology and education (4th ed.). New York: Wiley.

Unger, R. K., & Crawford, M. (1993). Commentary: Sex and gender - the troubled relationship between terms and concepts. Psychological Science, 4, 122-124.

Walkup, H., & Abbott, R. D. (1978). Cross-validation of item selection on the Bem Sex-Role Inventory. Applied Psychological Measurement, 2, 63-71.

Ward, C., & Sethi, R. R. (1986). Cross-cultural validation of the Bem Sex-Role Inventory. Journal of Cross-Cultural Psychology, 17, 300-314.

Webster, H. (1956). Personality development during the college years: Some quantitative results. Journal of Social Issues, 12, 29-43.

Welter, B. (1978). The cult of true womanhood, 1820-1860. In M.

Gordon (Ed.), The American family in social historical perspective (2nd ed.) (pp. 313-333). New York: St. Martin's Press.

Wheeless, V. E., & Dierks-Stewart, K. (1981). The psychometric properties of the Bem Sex-Role Inventory: Questions concerning reliability and validity. Communication Quarterly, 29, 173-186.

Wong, F. Y., McCreary, D. R., & Duffy, K. G. (1990). A further validation of the Bem Sex-Role Inventory: A multitrait-multimethod study. Sex Roles, 22, 249-259.

Worell, J., & Remer, P. (1992). Feminist perspectives in therapy: An empowerment model for women. New York: Wiley.

APPENDIX A

HOFFMAN GENDER SCALE (FORM A) (REVISED)

Hoffman Gender Scale (Form A)

PLEASE NOTE: Complete Form A if you are a female. Complete Form B (reverse side) if you are a male.

Please indicate your level of agreement with each of the following statements by rating it a "1," "2," "3," "4," "5," or "6" as follows:

1	2	3	4	5	6
Strongly Agree	**Agree**	**Tend to Agree**	**Somewhat Agree**	**Disagree**	**Strongly Disagree**

1. When I am asked to describe myself, being female is one of the first things I think of. ____
2. I am confident in my femininity. ____
3. I meet my personal standards for femininity. ____
4. My perception of myself is positively associated with my biological sex. ____
5. I am secure in my femininity. ____
6. I define myself largely in terms of my femininity. ____
7. My identity is strongly tied to my femininity. ____
8. I have a high regard for myself as a female. ____
9. Being a female is a critical part of how I view myself. ____
10. I am happy with myself as a female. ____
11. I am very comfortable being a female. ____
12. Femininity is an important aspect of my self-concept. ____
13. My sense of myself as a female is positive. ____
14. Being a female contributes a great deal to my sense of confidence. ____

What do *you* mean by femininity?

_____

_____

APPENDIX B

HOFFMAN GENDER SCALE (FORM B) (REVISED)

Hoffman Gender Scale (Form B)

PLEASE NOTE: Complete Form B if you are a male. Complete Form A (reverse side) if you are a female.

Please indicate your level of agreement with each of the following statements by rating it a "1," "2," "3," "4," "5," or "6" as follows:

1	2	3	4	5	6
Strongly Agree	**Agree**	**Tend to Agree**	**Somewhat Agree**	**Disagree**	**Strongly Disagree**

1. When I am asked to describe myself, being male is one of the first things I think of. ____
2. I am confident in my masculinity. ____
3. I meet my personal standards for masculinity. ____
4. My perception of myself is positively associated with my biological sex. ____
5. I am secure in my masculinity. ____
6. I define myself largely in terms of my masculinity. ____
7. My identity is strongly tied to my masculinity. ____
8. I have a high regard for myself as a male. ____
9. Being a male is a critical part of how I view myself. ____
10. I am happy with myself as a male. ____
11. I am very comfortable being a male. ____
12. Masculinity is an important aspect of my self-concept. ____
13. My sense of myself as a male is positive. ____
14. Being a male contributes a great deal to my sense of confidence. ____

What do *you* mean by masculinity?

_____

_____

APPENDIX C

INSTRUCTIONS TO PARTICIPANTS

This is a study designed to explore definitions of masculinity and femininity. I am interested in how you think about these terms, and what associations these words have for you. Should you choose to participate, you will be completing three instruments. I will go over the instructions for each instrument with you when it is time to complete each of them. Please do not read any of the instruments until I ask you to do so.

You do not need to write your name or identification number on any of the instruments; in fact, I ask that you do not provide this information, even though there are spaces for your name, etc., on the first instrument. You have the option to decide at any point during these activities that you do not wish to participate. If that is the case, you are asked to remain seated while others, who choose to, complete their assessments. If you decide not to participate, you may scribble or write on the test and it will be collected with the others. Are there any questions?

I will now distribute the packets. Please wait until everyone has received one before we proceed.

Please open the packet and take out the first instrument marked "# 1." I will go over the instructions for it with you. Please do not begin until the instructions have been read.

Read Instructions from BSRI.

When you are through please put the instrument back in the packet. Then wait for the next instructions.

Are there any questions?.... You may begin.

Wait until everyone has completed the BSRI and returned it to the envelope.

Now take out the instrument marked "Hoffman Gender Scale." You will notice that one side of the paper is described as Form A and the other side is described as Form B. Please make sure that you complete Form A if you are a female and Form B if you are a male. I will go over the instructions with you. Please do not begin until the instructions have been read.

Read instructions from HGS.

You will note that there are four levels of "agree" listed across the scale, and only two levels of "disagree." There is no "neutral" or "uncertain." Please read each question carefully. Many are very similar, but it is your responses to the subtle differences in meaning that are important.

When you are through please put the second scale back in the packet, and wait for further instructions.

Wait until everyone has completed the HGS and returned it to the packet.

The final task is to complete one additional instrument. Please take out the final instrument labeled "#3." You will note that it is a listing of the same words from the first instrument that you completed. I will go over the instructions with you. Please do not begin until the instructions have been read.

Read instructions from third instrument.

When you are through please make sure that you complete the items for age, class in school, and race on the second page. Then place the instrument back in the packet with the others and close the clasp on the packet. It is important that you do not review your responses to the first assignment, so please leave the completed instruments in the packet. Please remain seated while everyone else finishes and all packets have been collected. Thank you for your cooperation.

Are there any questions?.... You may begin.

Wait until everyone has completed the final instrument and returned it to the envelope. Collect the packets.