

TWENTY-FIVE YEARS AFTER THE BEM SEX-ROLE INVENTORY: A REASSESSMENT AND NEW ISSUES REGARDING CLASSIFICATION VARIABILITY

By: Rose Marie Hoffman and L. DiAnne Borders

Hoffman, R. M., & [Borders, L. D.](#) (2001). Twenty-five years after the Bem Sex-Role Inventory: A reassessment and new issues regarding classification variability. *Measurement and Evaluation in Counseling and Development*, 34, 39-55.

Made available courtesy of SAGE Publications (UK and US): <http://mec.sagepub.com/>

*****Note: Figures may be missing from this format of the document**

Abstract:

Respondents' Bem Sex-Role Inventory (BSRI; S. L. Bem, 1974) classifications may differ considerably on the basis of the form and scoring method used. The BSRI was reexamined with respect to past and present relevance.

Article:

The 1970s heralded a new concept in masculinity and femininity research: the idea that healthy women and men 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 was 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 a combination of stereotypically "feminine" and stereotypically "masculine" personality traits. The Bem Sex-Role Inventory (BSRI; Bem, 1974) was designed to facilitate empirical research on psychological androgyny. For the past quarter of a century, the BSRI has endured as the instrument of choice among researchers investigating gender role orientation (Beere, 1990).

Since its development in 1974, the BSRI has been widely used but also widely criticized. Ironically, early criticisms of the BSRI have contributed to its becoming even more well known as a masculinity-femininity measure and, consequently, used even more by researchers. In fact, it seems that the BSRI has been repeatedly used without sufficient attention to its theoretical framework (Frable, 1989); without clear and deliberate thought to the research questions being studied (Gilbert, 1985); and, as we argue, perhaps also without as thorough an understanding of the instrument as would be advisable.

It is interesting and salient that, after 25 years, Bem (1998) disclosed in her autobiography that she was not adequately prepared to develop this instrument and has been shocked by how popular it became and remains today. This honest admission clearly helps to explain some of the issues regarding this widely used instrument. The purpose of this article is threefold. The primary purpose is to explore the extent of variability among respondents' BSRI classifications (i.e., feminine, masculine, androgynous, and undifferentiated) depending on which form of the instrument (Original or Short) and which of the two scoring methods are used (i.e., median-split or hybrid [this latter scoring method uses both the median-split and the individual's Femininity- minus- Masculinity scores]). Both scoring methods are described in the test manual (Bem, 1981a) and are equally recommended by Bem. Classification variability has not been examined in previous research, a surprising observation given both the degree of attention that the BSRI has received since its inception and the emphasis researchers place on these classification categories in interpreting results of studies using this instrument. The second purpose is to reexamine 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. Although a study was conducted by Ballard-Reisch and Elton (1992) with this intent, a noncollege sample was used instead of college undergraduates, the group that Bem used to develop the BSRI.

The third purpose is to review and discuss theoretical and methodological issues related to the BSRI as the instrument marks 25 years as the most widely used measure in all areas of gender research. (A literature search conducted by Beere, 1990, in preparation for her anthology of gender tests and measures identified 795 articles and 167 ERIC documents that used the BSRI, and none of those references was a duplicate of those listed in her first book [Beere, 1979].) This article differs from previous critiques of the BSRI in that it addresses both the theoretical underpinnings of the BSRI and the key methodological issues in the development of the instrument and offers new perspectives as well as evaluates past criticisms. In the past, scholars tended to focus either on the conceptual framework for androgyny measures in general, rather than for the BSRI specifically (e.g., Ashmore, 1990; Frable, 1989; Lenney, 1991), or on particular methodological issues (e.g., factor structure, fourfold classification system), sometimes discussed specifically in relation to the BSRI (e.g., Blanchard-Fields, Suhrer-Roussel, & Hertzog, 1994) and other times examined in the broader context of androgyny literature (Marsh & Myers, 1986; Yamold, 1990). An exception is Pedhazur and Tetenbaum's (1979) critique, which identified several fundamental problems with the instrument early on. As we already suggested, however, many researchers who use the BSRI are not as knowledgeable about it as they may need to be. Thus, even criticisms that have been expressed previously may merit repeating. Moreover, Pedhazur and Tetenbaum's critique was written prior to the publication of the BSRI test manual in 1981, precluding any comments that Pedhazur and Tetenbaum could offer on the document that has guided BSRI use for the past two decades.

DEVELOPMENT OF THE BSRI

The BSRI differed from earlier instruments in that its developer, Sandra Bem, 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.

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 (1981b, 1981c) gender schema theory. 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. (Bem, 1981b, 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 (1979) 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" (p. 1048). Thus, she defined masculinity and femininity in terms of sex-linked social desirability.

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., truthful, conscientious, conceited). The 20 neutral items were used to constitute a measure of Social Desirability in response. Unlike the feminine items and the masculine items, all of which were 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).

To fully understand this instrument, one must be aware of the evolution of its scoring procedures. When the BSRI was first published in 1974, scoring 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 was true, the individual would be classified as "masculine"; and if the difference was small and not statistically significant, that person would be classified as "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 in scoring that resulted in her current procedure of using a median-split to form four distinct groups: feminine, masculine, androgynous, and undifferentiated. A difference score (between Femininity and Masculinity) is determined on the basis of standardized T-scores. 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 their sex).

A less popular scoring method, but one also presented in the manual (Bem, 1981 a), is the hybrid method, which differs from the median-split method in that it uses both the median-split and the difference between an individual's Femininity and Masculinity scores as bases for classification (see Bem, 1981 a, for a detailed description of this method). According to Bem (1981 a), "both [methods] appear to be perfectly adequate for research, but the Median-Split method is simpler to execute and explain" (p. 65). The comparative simplicity of the median-split method more than likely accounts for its overwhelming preference among researchers. It remains interesting, however, that given the extent of the critiques that the BSRI has prompted over the past 25 years, we found no studies that compared the hybrid and median-split scoring methods.

Soon after the development of the original version of the BSRI, Bem (1979, 1981 a) constructed the BSRI Short Form. It contains 30 of the original 60 items, with 10 items constituting each of the three scales (Masculinity, Femininity, and 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). Although the Short Form is generally viewed as more psychometrically sound than the Original Form (cf. Lippa, 1985; Payne, 1985), Bem (1981a) "strongly recommend[ed] the continued use of the original 60-item inventory" (p. 32) because of her belief that it predicted behavior better than the Short Form. Regardless, the standard instrument still lists all 60 items (the first 30 constitute the Short Form), and thus, the Original Form remains widely used despite the controversy regarding its value.

Psychometric issues aside, the BSRI and gender schema theory spurred notable changes in the way femininity and masculinity were conceptualized. For the first time, the cultural context was considered rather than an exclusive focus on differences in responses between the sexes to determine what was feminine and what was masculine. Moreover, Bem's work redeemed the relationship between psychological health and gender. The assumption that it was healthy for individuals to be sex-typed was replaced by her assertion that a combination of traditionally feminine and traditionally masculine qualities could be healthy regardless of one's biological sex. Sex differences were de-emphasized; however, the words masculine and feminine became increasingly popular as labels for specific characteristics.

CRITIQUES OF THE BSRI

The plethora of androgyny research using the BSRI has yielded many inconsistent findings and failures to replicate (see Ashmore, 1990; Cook, 1985). The BSRI is here examined in relation to its theoretical basis, validity, and factor analysis/dimensionality; item selection procedures; score interpretation; and reliability.

Theoretical Rationale: Gender Schema Theory

Bem did not define gender schema theory as such until several years after she developed the BSRI. Bem (1981c) referred to the process by which society "transmutes male and female into masculine and feminine" as "sex-typing" (p. 354) and contended that her theory explained why sex-typed individuals process information in gender-linked terms and nonsex-typed persons do not. We would argue, however, that one does not have to be sex-typed to use gender as a primary organizing principle. For example, women who identify as feminists are attuned to power imbalances resulting from gender inequities and thus attend closely to issues of gender, yet they are unlikely to be labeled feminine (i.e., sex-typed) by the BSRI classification system.

Bem seemed to conceptualize an individual as a "passive recipient of societal forces" (Ashmore, 1990, p. 507) rather than a complex being who participates in social constructions of gender. Furthermore, her theory depicts 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 prone to interpret information in masculine, but not feminine, terms or in feminine, but not masculine, terms (Markus, Crane, Bernstein, & Siladi, 1982).

Even if "cultural definition[s] of maleness and femaleness that constitute[d] the society's gender schema" (Bem, 1981b, p. 369) did exist, which is debatable given the variability within a culture (e.g., American society), it has been argued repeatedly that "maleness" and "femaleness" are quite different from the more limited concept of traditional male and female roles (Hoffman, Borders, & Hattie, in press; Spence, 1985). Furthermore, Bem (1985) later reconsidered the concept of androgyny and contended that "human behaviors should no longer be linked with gender, and society should stop projecting gender into situations irrelevant to genitalia" (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 on both sexes. However, Bem's earlier work provided the basis for the BSRI, which remains unchanged and widely used today. Given the discrepancies between Bem's more recent conceptualizations of gender and those that accompanied the development of the BSRI, one cannot refrain from questioning the implications of both past and present usage of the instrument.

Exactly what is being measured by the BSRI has long been questioned. Spence and Helmreich's (1981) analysis of the BSRI led them to conclude that, like their own instrument (i.e., Personal Attributes Questionnaire; Spence, Helmreich, & Stapp, 1974), the BSRI is basically a measure of instrumentality and expressiveness. Bem (1981b) objected, maintaining that instrumentality and expressiveness are only aspects of what the BSRI measures and that the instrument is a measure of masculinity and femininity as well as the individual's propensity to view the world using gender as a lens.

Validity of scores from the BSRI. Construct validity issues surrounding the BSRI result from the fact that femininity and masculinity remain inadequately and inconsistently defined in Bem's discussions. The construct validity of scores from the BSRI is further challenged (Lippa, 1985; Payne, 1985; Spence, 1984, 1985, 1991), because of the inconsistency in Bem's explanations of what the BSRI is intended to measure (i.e., one's global masculinity and femininity vs. one's tendency to use gender as a lens to view the world vs. gender-stereotypical or nonstereotypical self-descriptions). Given that construct validity supercedes all other types of validity (Messick, 1989), we argue that this inconsistency is a critical point.

In the test manual, Bem (1981a) described a series of validity studies (Bem, 1975; Bem & Lenney, 1976; Bem, Martyna, & Watson, 1976) designed 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 sex-typed (nonandrogynous) sex- role classification would demonstrate a more limited range of behavior across a variety of situations. For example, participants 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 validity of scores from the BSRI on the basis of the results of studies dealing with instrumental and expressive functioning. This

research consisted of four laboratory studies described in two articles (Bem, 1975; Bem et al., 1976). The first study was designed to measure independence under pressure to conform. The purpose of the other three studies was to measure nurturance with a kitten, a baby, and a lonely student, respectively. For men, results were clearer than for women. Men classified as "feminine" tended not to demonstrate independence and "masculine" men tended not to demonstrate nurturance. Although "feminine" women were low in independence and "masculine" women were low in nurturance, as demonstrated in their behavior with the baby and the lonely student, "masculine" women did display nurturance with the kitten. Androgynous persons of both sexes demonstrated independence and nurturance, depending on the situation. We observed that the manual does not include Bem's finding that "feminine" women did not display nurturance as predicted, although it is discussed in the report of the actual study (Bem et al., 1976). Bem (1987) contended that "taken as a whole, [these studies] 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). Others (Brannon, 1978; Lenney, 1991) supported Bem's view that there is sufficient evidence for the construct validity of scores from the BSRI. However, 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). Moreover, we argue that Lenney's (1991) qualification that the validity is adequate "when it is used in ways suggested by the theoretical rationale underlying its development" (p. 596) gives pause in light of the problems with the BSRI's theoretical rationale that we and others have identified.

An investigation of both the content and the process validity of BSRI scores conducted by Myers and Gonda (1982) failed to provide support for either type of validity. It is not surprising that one of their major criticisms focused on ambiguities in the definitions of masculinity and femininity. With respect to the process validity of BSRI scores, Myers and Gonda 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 positions presented by Lewin (1984), Spence (1985), and Hoffman (in press), who independently argued for the need to allow individuals their own personal definitions of masculinity and femininity.

Numerous validity studies have been conducted on the BSRI. It is the interpretation of most of those studies that remains the issue. Whether the BSRI successfully discriminates between individuals who adhere to sex role stereotypes and those who do not, construct validity cannot be adequately assessed as long as there are inconsistencies in Bem's accounts of what the BSRI is intended to measure.

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. Factor analyses typically depict two highly correlated instrumentality factors, one of which can be labeled dominance and the other self-reliance; an expressiveness factor; and a fourth factor defined by three BSRI items: feminine, masculine, and athletic (Lippa, 1985). In a similar manner, Pedhazur and Tetenbaum 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 composed 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 men's self-reports) as well as orthogonal to the other factors. More recently, Blanchard-Fields et al. (1994) conducted confirmatory factor analyses that supported a multidimensional factor structure.

Item Selection Procedures and Assessment of BSRI Items as Masculine, Feminine, or Neutral

In developing the BSRI item selection procedures, Bem (1987) sought to assess "what [was] collectively believe[d] to be the prevailing definitions of masculinity and femininity in the culture at large" (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). Again, however, confusion resulting from Bem's (1981c) inconsistent use of the terms masculinity and femininity as constructs measured by the BSRI is evident. On the other hand, Bem allows the individual to have personal definitions of

masculinity and femininity, yet she holds these definitions as irrelevant to gender-schematic processing and sex-typing (Ashmore, 1990). It does seem odd that a measurement tool composed of items selected for their sex-specific desirability can be used to validate a theory that 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, 1984, 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 approximately 400 items in her pool of adjectives was significantly (p less than .05) more desirable for a man than for a woman (to qualify as "masculine") or for a woman than for a man (to qualify as "feminine"). As Pedhazur and Tetenbaum noted, 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" when these procedures were used seems to support this observation and seems to exemplify the phenomenon of statistically significant results obscuring substantively meaningful findings. Furthermore, Bem failed to clarify in either her article that describes the development of the BSRI (1974) or the manual (1981a) exactly how many items met her selection criteria and how these were narrowed to 20 items for each of the two scales.

The judges of the social desirability of potential BSRI items (100 undergraduate students at Stanford University in 1972) 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, 1981 a, p. 11). Bem seemed to suggest that this procedure was used in an attempt to strengthen test construction; however, we argue that the result may have been the opposite: Bem's procedures provided no way of comparing how a judge would rate the desirability of an item for a woman versus a man.

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 (e.g., Edwards & Ashworth, 1977; Harris, 1994; Heerboth & Ramanaiah, 1985; Walkup & Abbott, 1978; Ward & Sethi, 1986) has been to assess the quality of BSRI items in terms of their identification as "masculine," "feminine," or "neutral." Results have been mixed.

Findings of at least one of the studies that supported the legitimacy of the BSRI items must be viewed cautiously. Our scrutiny of Harris's (1994) study revealed that the sample size was so large ($N = 3,000$) that significance was virtually inevitable. Moreover, although Harris described his work as "a replication study of item selection for the Bem Sex Role Inventory" (p. 241), he asked participants to evaluate (in terms of being "masculine" or "feminine") only the items that were ultimately included on the BSRI rather than the original total pool of adjectives.

Ballard-Reisch and Elton (1992) assessed a predominantly middle-class, Caucasian, noncollege population in the western United States for their interpretations of whether the 60 BSRI items were viewed as "masculine," "feminine," or "neutral." They found that 19 of the 60 items were viewed as neutral, only 1 item (i.e., feminine) was viewed as feminine, and only 1 item (i.e., masculine) was viewed as masculine. Agreement among participants in Ballard-Reisch and Elton's study did not reach the established agreement level (75%) with regard to the remaining 39 items. It is important to note, however, that unlike Harris (1994), Ballard-Reisch and Elton did not describe their study as an item selection replication but rather as a reexamination of the androgyny construct and its measurement. More recently, Spence and Buckner (2000) reported that gender stereotypes were supported when participants in their study were asked to compare and rate the "typical" male and female college student on Personal Attributes Questionnaire and BSRI instrumental and expressive items; however, such instructions seem to invite a tendency to stereotype by suggesting that a "typical" man or woman exists. As the preceding brief review suggests, results of studies of BSRI item selection as well as BSRI item evaluation are inconsistent.

Scoring of the BSRI

As we noted earlier, Bem (1977) recommended the use of a median-split technique, based on the criticism of Spence et al. (1975) that her original procedure did not discriminate between individuals who scored low on both the Masculine and Feminine scales and those who scored high on both scales. Before Bem's introduction of the median-split into BSRI scoring procedures, this method was used by Spence et al. (1974,1975) in computing scores on their instrument, the Personal Attributes Questionnaire. However, Spence and Helmreich (1978) voiced more concern than Bem regarding the use of this technique, in that it results in data subject to statistical distortion. They stressed that results obtained using this method should be interpreted with caution, particularly when research questions deal with between-group comparisons. 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 . . . constitute an additional source of 'noise' or error in any research design" (p. 9). However, our observations that (a) a considerable number of people often score near the cutoff points and (b) researchers seem to have consistently attached a considerable degree of importance to the classification of their participants according to the BSRI, as have many participants themselves, suggest that Bem's perspective is a serious minimization.

Although some researchers (e.g., Bryan, Coleman, & Ganong, 1981; Motowidlo, 1981; Orlofsky, Aslin, & Ginsburg, 1977) proposed alternative methods for scoring the BSRI and other androgyny measures, and other researchers (e.g., Myers & Sugar, 1979) further explored the relative merits of the median-split method and the alternative methods proposed by others than Bem, Bem did not respond to these suggestions for alternative scoring approaches, nor did she ever revise scoring procedures after the publication of the BSRI manual in 1981. The fact that the manual has not been revised in its 20 years of existence constitutes an additional area of concern, given that the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) suggest that test manuals be systematically revised to reflect additional evidence of reliability and validity. Regardless, the median-split and hybrid methods remain the only procedures that Bem recommended and were equally supported in her view. Moreover, although some researchers did explore alternatives to the median-split method, none considered the hybrid method, which was the only other scoring method that Bem recommended. The fourfold classification system by which the BSRI categorizes individuals as feminine, masculine, androgynous, or undifferentiated renders one's Masculinity and Femininity scale scores relatively meaningless in and of themselves. It is only in relation to the medians and to each other that the scale scores are important, because this is how one's classification is derived, and it is the classification of respondents as feminine, masculine, androgynous, or undifferentiated that forms the basis for the conclusions drawn by researchers using the BSRI.

Researchers wishing to use the BSRI rely on scoring procedures described in the BSRI manual included in the administration packet. As we already noted, these are the mediansplit and hybrid methods, both of which yield the fourfold classification system. Bem (1981a) noted in the manual that "the two methods which overlap a great deal . . . achieve perfect agreement for 76 percent of the subjects on the Original BSRI and for 71 percent of the subjects on the Short BSRI" (p. 64). In other words, however, differences in classification existed for 24% of Bem's respondents on the Original BSRI and for 29% of Bem's respondents on the Short Form of the instrument. Whether the degree of variability in Bem's test development sample is acceptable is certainly open to question; however, we found nothing about this in the literature. Considering the BSRI's popularity, it is surprising that our review revealed no studies that have examined the extent or ramifications of classification variability with respect to the BSRI Original Form versus the BSRI Short Form, in relation to the median-split versus the hybrid scoring method, or with respect to the combination of these two variables.

Reliability of Scores From the BSRI

Two samples, obtained in 1973 and 1978, provided the basis for reliability data reported by Bem (1981a). 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. Bem reported test-retest reliabilities

over a 4-week time span that ranged from .76, for males describing themselves on the Masculinity scale items (both Original and Short Forms), to .94, for females describing themselves on the Masculinity scale items (Original Form). Although much less controversy surrounds the reliability than the validity of BSRI scores, reliability without validity is of questionable value.

Summary

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). Yet the manual advises researchers to use the Original Form, and the standard instrument provided in the packet that researchers must purchase contains all 60 items. More fundamental issues, particularly those related to the BSRI's theoretical rationale and to item selection procedures, provide sufficient evidence to warrant considerable doubt regarding the use of the BSRI in research designed to assess masculinity and femininity. Moreover, research suggests that perceptions of femininity and masculinity in the 1990s and beyond are different from perceptions of these constructs in 1974 (Ballard-Reisch & Elton, 1992), thus challenging the meaningfulness of the BSRI as well as previous masculinity-femininity measures. Finally, the comparative BSRI classifications of an individual on the basis of the form of the instrument (i.e., Original or Short) and scoring method (i.e., median-split or hybrid) remains unexamined, despite the extensive research that the instrument has generated and despite the fact that BSRI classifications typically constitute the basis for the conclusions drawn by researchers who use the BSRI. The following study attempted to explore some of these issues.

A COMPARISON OF BSRI CLASSIFICATIONS BY FORM AND SCORING METHOD

The primary purpose of this study was to investigate the extent of variability of BSRI respondents' classifications related to form of the instrument (i.e., Original or Short) and scoring method (i.e., median-split or hybrid) used. A secondary aim was to reexamine the items that constitute the BSRI in terms of their viability as representations of current perceptions of masculinity and femininity among college undergraduates.

Method

Participants. Participants were 273 women and 98 men enrolled in 10 undergraduate classes in the departments of human development and family studies, public health education, and counseling at a moderately sized university in the southeastern United States. Two groups of student-athletes in classes in the university's Athletic Enhancement Program also participated. Age of respondents ranged from 17 to 46 years ($M = 20.45$, $SD = 4.12$, $Mdn = 19$). Most of the participants who identified their ethnicity were White/ Caucasian ($n = 244$). A total of 91 identified themselves as African American, Hispanic, Native American, Asian, or other. (Because the numbers in each individual ethnic minority category were inadequate for analysis, data were combined.) Ethnic identity was not reported by 36 participants. There were 132 first-year students, 84 sophomores, 70 juniors, and 49 seniors. Thirty-six participants did not report their year.

Procedure. As part of a larger study, all participants were first asked to complete the BSRI as a self-report. They were then asked to go through a listing of the BSRI items and rate each of the 60 items as feminine, masculine, or neutral. Rose Marie Hoffman administered the assessments to all 12 groups of participants. Students were advised that participation was voluntary and were given the option of using that time to complete other work. Three students declined participation.

Results

BSRI classifications by form and scoring method. Participants' self-descriptions on the BSRI were scored using two methods: (a) a median-split classification system based on 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).

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, and undifferentiated) was calculated and then compared with 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 5.39 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 5.00 for the present sample. Also, for this sample, the correlation between the Original and Short Form Masculinity scores was .92, and the correlation between the Original and Short Form Femininity scores was .89. Table 1 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. Data are reported separately for women and men.

The relationships between Bem's sample and the present sample were in marked contrast depending on the scoring method used. When the results obtained by the hybrid method were compared with those obtained by the median-split method, it should be recalled that Bem (1981 a) found that differences in classification occurred for approximately one fourth of respondents (24% on the Original Form and 29% on the 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 34% of respondents on the Short Form. For 24% of participants, classification differed by scoring method on both the Original and the Short Forms. Moreover, nearly one fifth (19%) of respondents were classified as three of the four categories (i.e., feminine, masculine, androgynous, and undifferentiated), depending on which of the four possible combinations of method and form was used. Six of the participants were classified in a different category for each of the four possible combinations of method and form and, thus, could be labeled by researchers as feminine, masculine, androgynous, and undifferentiated, depending on which form and method was used (see Table 2).

Perception of BSRI items as masculine, feminine, or neutral. The second set of analyses assessed the level of agreement among college undergraduates supporting the "masculinity" of the items that constitute the BSRI Masculinity scale and the "femininity" of the items that constitute the BSRI Femininity scale. As noted, this is a partial replication of Ballard-Reisch and Elton (1992), but with a sample that more closely resembled Bem's sample (i.e., college undergraduates). To assess level of agreement, the calculation of an index of neutrality level for each BSRI item was required. 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 is recognized as an indication of "extensive agreement" in research on stereotypes (Ballard-Reisch & Elton, 1992, p. 296; Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972, p. 62). Following the design used by both Bem (1981 a) and Ballard-Reisch and Elton (1992), male and female participants' responses (i.e., assessments of items as feminine, masculine, or neutral) were combined for analysis.

Of the 60 BSRI items, 22 items were determined to be neutral by at least 75% of the participants. Of these 22 items, 9 were from the BSRI Masculinity scale: defend my own beliefs (94%), ambitious (88%), individualistic (86%), self-reliant (85%), self-sufficient (84%), independent (83%), willing to take a stand (81%), strong personality (80%), and have leadership abilities (80%); 1 was from Bem's Femininity scale: loyal (75%); and the remaining 12 items were filler or neutral items: conscientious (77%), reliable (83%), truthful (85%), adaptable (82%), conventional (83%), helpful (78%), unsystematic (76%), inefficient (85%), happy (89%), solemn (81%), likable (91%), and friendly (89%). Masculine was the only 1 of the 60 BSRI items to reach a 75% agreement level to be classified as masculine, and feminine 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 levels of agreement for the present study can be found in Table 3.

Multivariate analyses of variance were conducted to examine the possible effects of race and year in school on BSRI classifications as well as on participants' assessments of BSRI items as feminine, masculine, or neutral. No relationship was found for either demographic variable.

Discussion

Variability of BSRI classifications by form and scoring method. The primary purpose of this study was to examine the variability in individuals' BSRI classifications (i.e., feminine, masculine, androgynous, and undifferentiated) depending on which form of the instrument is used (Original or Short) as well as on which of Bem's (1981a) two recommended scoring methods is used (median-split or hybrid). Results suggest that such variability can be extensive. It is particularly revealing that 41% of the respondents on the BSRI Original Form had two different classifications. Considering that the BSRI Original Form is the one advocated by Bem (1981a) in the manual, and that the instrument provided in the test packet that researchers must purchase is the Original Form, this is cause for some concern. These findings are especially informative because of absence of attention to this phenomenon in previous research.

Future research should be conducted to examine the degree of such variability among other samples. However, it seems that, regardless of outcomes of subsequent studies of the extent to which respondents' classifications vary in relation to BSRI form, scoring method, or both, the acceptability of the extent of such differences reported by Bem (1981a) on her sample (24% on the Original Form and 29% on the Short Form) is indeed questionable.

It should be emphasized that it is the respondent's classification as feminine, masculine, androgynous, or undifferentiated, rather than individual Masculinity and Femininity scale scores, that provides the basis for most researchers' conclusions using the BSRI. If respondents' classifications fluctuate easily on the basis of form and scoring method, as the results of this study suggest, then conclusions of research based on BSRI classifications clearly must be reevaluated.

Perceptions of femininity and masculinity. A secondary purpose of this study was to assess contemporary college undergraduates' perceptions of femininity and masculinity using the BSRI items as a vehicle. Consistent with the findings of Ballard-Reisch and Elton (1992), masculine and feminine were the only items on the entire inventory that met the 75% agreement level necessary to be classified as such. The remaining 19 items on the BSRI Masculine scale and the remaining 19 BSRI Feminine scale items failed to meet this criterion. Clearly, college undergraduates in this study perceived BSRI items very differently from the gender-stereotypical way that the 1970s college undergraduates who served as judges in Bem's test development process viewed these descriptors.

These differences give further cause to doubt the current theoretical meaningfulness of the fourfold classification system by which BSRI scale scores are usually interpreted. If the items that constitute the BSRI Masculinity scale are no longer considered masculine and the items on the BSRI Femininity scale are no longer considered feminine, then the basis for classifying individuals in such terms is eroded. The results of this study suggest that gender schema theory (Bem, 1981a), which relied on cultural definitions of masculinity and femininity as a framework for one's organization of information about self and others, may be less relevant than before.

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). Androgyny suggested that individuals could exhibit both masculine and feminine traits. The findings previously described suggest that traits are no longer perceived in those terms. Therefore, these findings, in conjunction with those of Ballard-Reisch and Elton (1992), suggest that androgyny may indeed be less useful a concept than it once was, as our conceptualizations of femininity and masculinity within the broader culture undergo change.

CONCLUSION

The assumption underlying the development of the BSRI, that so-called masculine traits and feminine traits are mutually exclusive (e.g., if independent is considered "masculine" by a man in his definition of self, then it cannot be considered "feminine" by a woman in her definition of self), is an erroneous assumption (Lewin, 1984), whether the setting is the 1970s or the twenty-first century. This assumption underlies the conceptual and methodological issues with the BSRI discussed throughout this article. An even more fundamental problem relates to the question of what is being measured by the BSRI. Instrumentality and expressiveness do not equal masculinity and femininity.

That there are problems regarding the BSRI and its use seems clear. 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.

Although the BSRI has undoubtedly been useful as an impetus for research and discussion related to gender role identity and other gender-related constructs (gender identity, gender role salience, etc.), and its developer has been unequalled 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. We argue that it is no longer useful to conceptualize traits as such, given the results of this study and of Ballard-Reisch and Elton (1992), both of which indicate that masculine and feminine are the only BSRI items currently perceived as "masculine" and "feminine."

Androgyny was proposed as a means to ameliorate some of the problems historically associated with the study of masculinity and femininity (Cook, 1987). However, as Lott (1981) argued, androgyny theory's continued reliance on traditional notions of femininity and masculinity served to reify the very distinction that it sought to blur. In fact, the very labeling 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).

The naming of the two BSRI scales as Masculine and Feminine certainly was not intended to cement gender-stereotypical thinking. On the contrary, Bem's purpose was to stimulate research on androgyny, and this she did. But, as an extensive feminist literature suggests (cf. Enns, 1993; Kaschak, 1992), naming is a powerful phenomenon that, in this example, can serve to maintain rather than ameliorate a dichotomy and, with it, a status quo.

The results of this study, in conjunction with the previous work of many scholars, suggest that the usefulness and meaningfulness of the BSRI, both present and past, are indeed debatable. Researchers are encouraged to carefully consider the relevance of classifying participants according to a system that is both flawed and outmoded. The BSRI may best be used to examine one's instrumentality and expressiveness, providing that the importance attached to one's classification is minimized and providing that instrumentality and expressiveness are described as human traits rather than as synonyms for masculinity and femininity. Researchers also might explore new models for conceptualizing femininity and masculinity in nonstereotypical ways (Hoffman et al., in press).

Theoretical and psychometric problems notwithstanding, the BSRI has served for 25 years as a vehicle for empirical research in masculinity, femininity, and androgyny. It is certainly largely to Bem's credit that we have been challenged to think critically about such constructs. Nevertheless, it is now time to build on her work by ceasing to reinforce the dichotomy between women and men and by beginning to more fully explore the possibilities of the type of society that Bem has supported in her writing.

TABLE 1

Percentages of Participants by Category in BSRI Fourfold Classification System
 Legend for Chart:

A - Sample

B - Feminine

C - Masculine

D - Androgynous

E - Undifferentiated

A	B	C	D	E
---	---	---	---	---

BSRI Development Sample

Women-Original Form

Median-split method	39.4	12.4	30.3	17.9
Hybrid method	41.2	10.0	24.1	24.7

Women-Short Form

Median-split method	23.8	15.6	37.1	23.5
Hybrid method	29.4	16.2	27.4	27.1

Men-Original Form

Median-split method	11.6	42.0	19.5	26.9
Hybrid method	12.2	40.8	14.1	33.0

Men-Short Form

Median-split method	16.0	32.6	23.9	27.5
Hybrid method	14.7	28.2	19.1	38.0

Current Study Sample

Women-Original Form

Median-split method	34.4	17.6	25.6	22.3
Hybrid method	46.2	8.8	33.3	11.7

Women-Short Form

Median-split method	42.9	9.9	33.3	13.2
Hybrid method	37.4	10.6	45.4	5.9

Men-Original Form

Median-split method	4.1	55.1	22.4	18.4
Hybrid method	10.2	45.9	26.5	17.3

Men-Short Form

Median-split method	14.3	29.6	33.7	21.4
Hybrid method	9.2	35.7	49.0	5.1

Note. BSRI = Bem Sex-Role Inventory.

TABLE 2

Frequency and Percentage of Participants Having 1, 2, 3, or 4 BSRI Classifications

Description of Participant	n	% of Sample
----------------------------	---	-------------

Had two different scores on Original Form (for median-split vs. hybrid scoring)	151	41
---	-----	----

Had two different scores on Short Form (for median-split vs. hybrid scoring)	125	34
--	-----	----

Had two different scores on both forms (for median-split vs. hybrid scoring)	89	24
--	----	----

Were in 3 out of 4 possible classifications (across form and scoring method)	72	19
--	----	----

Were in 4 out of 4 possible classifications (across form and scoring method)	6	2
--	---	---

Note. See Table 1 Note. N = 371.

TABLE 3

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

Legend for Chart:

- A - BSRI Item
- B - Feminine: n
- C - Feminine: %
- D - Masculine: n
- E - Masculine: %
- F - Neutral: n
- G - Neutral: %

A	B	C	D		
	E	F	G		
1. Defend my own beliefs (M)	4	349	94	8	2 14
2. Affectionate (F)	0	158	57	43	0
3. Conscientious (N)	4	69	77	19	16

4. Independent (M)	13	17	5	48
		306	83	
5. Sympathetic (F)	0	201	54	0
		170	46	
6. Moody (N)	6	132	36	21
		218	59	
7. Assertive (M)	36	12	3	132
		227	61	
8. Sensitive to others' needs (F)	0	207	56	0
		164	44	
9. Reliable (N)	5	43	12	19
		309	83	
10. Strong personality (M)	17	11	3	64
		296	80	
11. Understanding (F)	0	140	38	1
		230	62	
12. Jealous (N)	15	44	12	57
		269	73	
13. Forceful (M)	66	0	0	244
		125	34	
14. Compassionate (F)	0	183	50	0
		187	51	
15. Truthful (N)	0	53	14	1
		317	85	
16. Have leadership abilities (M)	18	9	2	66
		296	80	
17. Eager to soothe feelings (F)	1	228	62	4
		139	38	
18. Secretive (N)	14	82	22	51
		237	64	
19. Willing to take risks (M)	43	1	0	160
		210	57	
20. Warm (F)	1	166	45	3
		202	54	
21. Adaptable (N)		35	10	32

	9	303	82	
22. Dominant (M)		1	0	204
	55	166	45	
23. Tender (F)		209	56	1
	0	161	43	
24. Conceited (N)		15	4	97
	26	257	70	
25. Willing to take a stand (M)		9	2	61
	16	301	81	
26. Love children (F)		109	29	0
	0	262	71	
27. Tactful (N)		65	18	40
	11	264	72	
28. Aggressive (M)		1	0	193
	52	177	48	
29. Gentle (F)		183	49	7
	2	181	49	
30. Conventional (N)		34	9	30
	8	304	83	
31. Self-reliant (M)		8	2	47
	13	314	85	
32. Yielding (F)		123	33	6
	2	239	65	
33. Helpful (N)		78	21	4
	1	287	78	
34. Athletic (M)		2	1	98
	27	269	73	
35. Cheerful (F)		106	29	2
	1	261	71	
36. Unsystematic (N)		21	6	66
	18	282	76	
37. Analytical (M)		37	10	63
	17	269	73	
38. Shy (F)		112	30	5
	1	252	68	

39. Inefficient (N)	11	15 314	4 85	40
40. Makes decisions easily (M)	24	257	23 70	6 89
41. Flatterable (F)	3	136 222	37 60	11
42. Theatrical (N)	5	97 255	26 69	17
43. Self-sufficient (M)	12	17 309	5 84	43
44. Loyal (F)	4	80 276	22 75	13
45. Happy (N)	1	37 328	10 89	4
46. Individualistic (M)	10	14 319	4 86	36
47. Soft-spoken (F)	1	225 142	61 39	2
48. Unpredictable (N)	21	39 253	11 69	77
49. Masculine (M)	78	1 81	0 22	287
50. Gullible (F)	4	131 223	36 60	15
51. Solemn (N)	11	30 300	8 81	39
52. Competitive (M)	36	1 235	0 64	133
53. Childlike (F)	20	50 245	14 66	74
54. Likable (N)	1	31 334	8 91	4
55. Ambitious (M)	10	10 323	3 88	36

56. Do not use harsh language (F)	155	42	6
	2	208	56
57. Sincere (N)	100	27	4
	1	265	72
58. Act as a leader (M)	4	1	99
	27	266	72
59. Feminine (F)	291	79	3
	1	75	20
60. Friendly (N)	41	11	1
	0	326	89

Note. See Table 1 Note. M = masculine; F = feminine; N = neutral.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- 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 Press.
- Ballard-Reisch, D., & Elton, M. (1992). Gender orientation and the Bem Sex Role Inventory: A psychological construct revisited. *Sex Roles*, 27, 291-306.
- Beere, C. A. (1979). *Women and women's issues: A handbook of tests and measures*. San Francisco: 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. (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, 369-371.
- 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, NE: 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. (1998). *An unconventional family*. 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.
- Blanchard-Fields, F., Suhrer-Roussel, L., & Hertzog, C. (1994). A confirmatory factor analysis of the Bem Sex Role Inventory: Old questions, new answers. *Sex Roles*, 30, 423-457.
- 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.
- Bryan, L., Coleman, M., & Ganong, L. (1981). Geometric mean as a continuous measure of androgyny. *Psychological Reports*, 48, 691-694.
- 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.
- 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. (1993). Twenty years of feminist counseling and therapy: From naming biases to implementing multifaceted practice. *The Counseling Psychologist*, 21, 3-87.
- Enns, C. Z. (1994). Archetypes and gender: Goddesses, warriors, and psychological health. *Journal of Counseling & Development*, 73, 127-133.
- Frale, D. E. S. (1989). Sex typing and gender ideology: Two facets of the individual's gender psychology that go together. *Journal of Personality and Social Psychology*, 56, 95-108.
- 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.
- 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.
- 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.
- Hoffman, R. M. (in press). The measurement of masculinity and femininity: An historical perspective with implications for counseling. *Journal of Counseling & Development*.
- Hoffman, R. M., Borders, L. D., & Hattie, J. A. (in press). Reconceptualizing femininity and masculinity: From gender roles to gender self-confidence. *Journal of Social Behavior and Personality*.
- Kaschak, E. (1992). *Engendered lives*. New York: Basic Books.
- 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.). (1984). 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.
- 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, NE: Buros Institute of Mental Measurements.
- 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.
- 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.
- 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.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education/Macmillan.
- Motowidlo, S. J. (1981). A scoring procedure for sex-role orientation based on profile similarity indices. *Educational and Psychological Measurement*, 41, 735-745.
- Myers, A. M., & Gonda, G. (1982). Empirical validation of the Bem Sex-Role Inventory. *Journal of Personality and Social Psychology*, 43, 304-318.
- Myers, A. M., & Sugar, J. (1979). A critical analysis of scoring the BSRI: Implications for conceptualization. *Catalog of Selected Documents in Psychology*, 9(2), (Ms. No. 1833, p. 24).
- Orlofsky, J. L., Aslin, A. L., & Ginsburg, S. D. (1977). Differential effectiveness of two classification procedures on the Bem Sex Role Inventory. *Journal of Personality Assessment*, 41, 414-416.
- 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, NE: Buros Institute of Mental Measurements.
- Pedhazur, E. J., & Tetenbaum, T. J. (1979). Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology*, 37, 996-1016.
- 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, NE: 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., & Buckner, C. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: What do they signify? *Psychology of Women Quarterly*, 24, 44-62.
- Spence, J. T., & Helmreich, R. L. (1978). *Masculinity & femininity: Their psychological dimensions, correlates, & antecedents*. Austin: 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.
- 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.
- Yarnold, P. (1990). Androgyny and sex-typing as continuous independent factors, and a glimpse of the future. *Multivariate Behavioral Research*, 25, 407-419.