Sex Roles, Vol. 30, Nos. 5/6, 1994

A Confirmatory Factor Analysis of the Bem Sex Role Inventory: Old Questions, New Answers[1]

Fredda Blanchard-Fields[2] and Lynda Suhrer-Roussel
Louisiana State University

Christopher Hertzog
Georgia Institute of Technology

LISREL analyses of data from a sample of 671 adults (90% Caucasian, 10% Black) evaluated (a) item factor structure of the Bem Sex Role Inventory (BSRI), (b) second-order factor model for the item factors, and (c) structural equation models estimating age and gender differences in these factors. Seven first-order item factors were extracted and found to have equivalent factor loadings for males and females. Item factors were related to two second-order factors: Masculinity and Femininity. There were relatively small age and gender differences in the first- and second-order factors. There was a differential relationship between self-rated masculinity and femininity and the first-order BSRI item factors for males and females. Results suggest that the BSRI best assesses gender-related personality traits and represents only one component of the complex multidimensional construct of gender roles.

The Bem Sex Role Inventory (BSRI) was originally developed as a self-report measure of the global constructs masculinity (M) and femininity (F) in order to identify gender-typed and nongender-typed (i.e., androgynous and cross-typed) individuals. The past 15 years has seen a proliferation of research questioning the conceptualization and measurement of gender role orientation (Ashmore, 1990; Ashmore & Del Boca, 1986; Bem, 1984;

[1]We thank Herbert W. Marsh and an anonymous reviewer for valuable comments on previous drafts of this article.
[2]To whom correspondence should be addressed at the School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170.

0360-0025/94/0300-0423$07.00/0

Pedhazur & Tetenbaum, 1979; Spence, 1984a). Recent work has supported Constantinople's (1973) challenge to the bipolarity assumption of masculinity and femininity. There is a general consensus in the theoretical and empirical literature that argues that masculinity and femininity should be treated as correlated, but distinct constructs (see Marsh & Myers, 1986, for a review). Constantinople (1973) also argued for the inherent multidimensionality underlying masculinity and femininity. Similarly, Ashmore (1990) argues that Bem's procedure for gender role classification on the basis of M and F personality constructs does not capture the multidimensionality of the gender identity construct. Instead, M and F can be conceptualized as a profile of scores on gender-related personality attributes (Spence, 1984a). If so, then it would appear that the bulk of the items in the BSRI reflect two well-known dimensions of personality in the interpersonal domain: dominance (agency) and nurturance (communion; (see Wiggins, 1979; Wiggins & Pincus, 1992). Indeed, one could argue that the BSRI estimates these two personality dimensions and merely relabels them as masculinity and femininity.

However, a considerable number of exploratory factor analyses of the BSRI items have suggested that two factors are inadequate for representing its factor structure. Although a few studies have argued for a two-factor (M and F) solution (Bledsoe, 1983; Carlsson, 1981; Thompson & Melancon, 1986), they can be criticized for inadequate evaluation of alternative, more differentiated factor structures. Most studies have identified at least four major BSRI item factors: a bipolar M-F dimension (defined by the Masculinity and Femininity items alone), Interpersonal Sensitivity, Assertiveness, and Self-Sufficiency (Collins, Waters, & Waters, 1979; Gross, Batlis, Small, & Erdwins, 1979; Larsen & Seidman, 1986; Pedhazur & Tetenbaum, 1979; Ruch, 1984; Schmitt & Millard, 1988; Waters & Popovich, 1986; Waters, Waters, & Pincus, 1977). This solution thus breaks the M dimension into two factors: Assertiveness and Self-Sufficiency. Other studies have found two to seven additional factors including further subdivision of masculinity items to include factors of Leadership and Athletic/Competitive (Gaa, Liberman, & Edwards, 1979; Maznah & Choo, 1986; Sassenrath & Yonge, 1979) and multiple femininity subfactors, including Introversion and a differentiation of items loading on Interpersonal Sensitivity (Berzins, Welling, & Wetter, 1978; Feldman, Biringen, & Nash, 1981; Hiller & Philliber, 1985; Windle & Sinnott, 1985).

Comparisons of factor solutions across these studies is hampered by a number of issues. A major source of variation in the published literature is the large diversity in types of samples studied, including the age range of respondents, gender of respondents, and whether samples of males and females were analyzed in aggregated or separate models. A second issue involves differences between studies in the variables actually used in the factor analysis. In many cases subsets of BSRI items were factored, or the BSRI items were factored in the presence of other variables (including the filler items from the BSRI). Factor solutions can and do change as a function of the selection of variables to be factored (Mulaik, 1972). The chief problem for comparing the existing factor analyses of the BSRI, however, is the virtually exclusive use of exploratory factor analytic (EFA) techniques, especially the typical practice of using principal components to extract factors, followed by a varimax rotation to an orthogonal solution. This approach is questionable given that it seems likely the attributes measured by the BSRI are correlated and not statistically orthogonal. Imposing orthogonal solutions on inherently correlated structures can result in erroneous inferences regarding identification of factors (e.g., Hertzog, 1989). Finally, exploratory methods provide no objective and defensible means for comparing factor structures across different populations—and so one cannot know from the existing literature whether the BSRI has an equivalent factor structure in males and females.

With mounting evidence for the multidimensionality of the BSRI and the existing diversity and limitations in methodology across studies, there is a need for studies that employ a more precise methodology such as confirmatory factor analysis (CFA) to test hypotheses of a lower and higher order factor structure of the BSRI. Marsh and Myers (1986) reviewed three separate CFA studies of gender role inventories, including the BSRI, conducted by Marsh and his colleagues. Their overarching model provides a useful backdrop for the theoretical rationale for the present study. They suggest that the global M and F factors should be viewed as separate but correlated higher order constructs defined by a number of specific traits. They present a possible representation of such a hierarchical model of global M and F by including specific traits suggested by previous research. Similar to the EFA research, Femininity traits include emotional, dependent, nurturant, traditional, empathetic/sensitive to others, verbally expressive, and romantic. Again, similar to previous EFA findings, Masculinity traits include aggressive/dominant, self-sufficient, competitive/assertive, physical/athletic, rational, goal directed, and tough. By positing such a model, it is possible to test related hypotheses using a CFA analysis including (a) how this structure is manifested in gender role inventories such as the BSRI, (b) the similarity of such a structure in multiple groups (i.e., gender groups), and (c) relationships between the global factors within the gender role inventory themselves as well as with biological gender, masculine and feminine self-ratings, age, etc.

Although Marsh and Myers (1986) posited this model based on their review of several CFA studies, their findings regarding the factor structure of the BSRI are of primary interest to this study. Marsh and Myers (1986) fitted a two-factor structure (M and F) to 14 feminine and 14 masculine BSRI items on a sample of 269 adolescent females. They found the M and F factors to be positively and significantly correlated (.58), indicating both that (a) M and F are correlated dimensions, rendering orthogonal solutions for the BSRI problematic, and (b) M and F are not polar opposites. However, their results were not definitive regarding the BSRI factor structure. Several items had nonsignificant loadings on the two factors, and the moderate goodness-of-fit of the two-factor model suggested that additional factors might have been present (even in this reduced item set).

Windle and Sinnott (1985) conducted a preliminary CFA on the BSRI in a sample of older adults that rejected the two-factor model for the entire 40-item set. They estimated factor correlations for M and F of .09 for older females and .32 for older males. Thus, this study also appears to be at odds with the hypothesis that M and F are uncorrelated, or even negatively correlated, factors.

The present study was designed to investigate the structure of the BSRI using CFA and structural equation models. We set out to investigate several issues regarding BSRI item structure. Our first goal was to identify an appropriate CFA model for the BSRI item factors, testing a number of alternative specifications. We hypothesized, based upon both the theoretical model of Marsh and Myers (1986) and the exploratory factor analytic literature, that a relatively large number of factors would be needed to fully account for the BSRI item correlations, although some of these factors could be substantially intercorrelated. Second, assuming that multiple correlated factors would be identified, we planned on using CFA to provide a direct test of the hypothesis that these intercorrelated item factors could be successfully modeled with a single higher order bipolar M-F factor, as in the study by Marsh (1985), or in terms of two higher order global M and F factors (as implied by the original conceptualization of Bem, 1974, and Marsh & Myers, 1986).

This study also addresses a third issue of the factorial invariance of the BSRI across gender groups. Much of the existing literature analyzed samples combining males and females, implicitly assuming gender equivalence in factor structure. Pedhazur and Tetenbaum (1979) analyzed gender groups separately, and reported four factors for both female and male subsamples that differed somewhat in structure. Windle and Sinnott (1985) extracted different numbers of factors in separate samples of male and female adults, and argued for qualitative gender differences in the underlying factor structure. Unfortunately, neither study attempted factor matching rotations (e.g., Meredith, 1964), and hence it is possible that differences in factor structures they obtained were an artifact of computing separate exploratory factor analyses in each gender group (see Meredith, 1964; Mulaik, 1972). Indeed, Windle (1986) reported a subsequent simultaneous CFA of a subset of 10 BSRI items from Windle and Sinnott's (1985) data. He found equivalent factor loadings across gender groups. The present study produced a more comprehensive evaluation of the hypothesis of factorial invariance across gender using CFA.

A fourth issue that needs to be examined using a structural equation modeling framework is the relationship between gender and the M and F constructs. Research has typically found the expected pattern of gender-related correlations with gender role traits, i.e., females score higher on femininity factors and males score higher on masculinity factors, although the proper interpretation of these differences remains a matter of debate (Marsh & Myers, 1986). Although statistically significant gender differences in BSRI scales (and individual items) are typically found (e.g., Feldman et al., 1981), these differences may not be particularly large when evaluated in terms of effect size. At the same time Masculine and Feminine items show a strong relationship to gender. Thus, it is not surprising that the use of the Masculine and Feminine items in the M and F scales has a major impact on the magnitude of gender differences. Pedhazur and Tetenbaum (1979) found that the bipolar factor formed by the Masculine and Feminine items accounted for most of the gender-related variance in the BSRI, a finding that led Bem (1979) to recommend dropping these items from a revised version of the scale. Marsh and Myers (1986) argued that the bipolar factor substantially correlates with both global F and global M, suggesting that it may actually mediate the relationship between gender and these more global constructs. In any event, it seemed important to evaluate gender differences on the BSRI item factors computed without the Masculine and Feminine item self-ratings. The advantage of structural equation models in this context is that the item factors can be regressed directly on gender in the model, estimating the regression coefficients while avoiding the methodological issues that arise in estimating the correlations using summated scales or factor scores from BSRI items.

The evidence favoring differentiation of the M-F item bipolar factor from the global M and F constructs raises the fifth and final issue: how should this bipolar factor be conceptualized? Perhaps, as suggested by Ashmore (1990), these ratings are reflections of a component of gender identity, whereas other BSRI factors reflect gender-correlated aspects of personality (Spence, 1984a). If so, then the fact that the Masculine and Feminine items still form a salient bipolar factor when gender groups are factored separately suggests that within-gender individual differences in self-rated masculinity and femininity may have differential relationships to the other attributes measured in gender role inventories. The possibility that gender itself will moderate the relationship between self-rated masculinity and femininity and BSRI factors has, to our knowledge, not been investigated. It could well be the case, for example, that males and females exhibit similar organization and structure of gender-related attributes, but differ in the extent to which self-ratings of masculinity and femininity correlate with these attributes. We used the results from CFA of each gender group to test formally the hypothesis of gender differences in relationships of masculinity and femininity ratings to BSRI item factors.

METHOD

Participants

The sample consisted of 723 individuals representing five age groups from the greater Baton Rouge, Louisiana, area. The adult participants (90% Caucasian, 10% Black) were unpaid volunteers from the community. The youngest adults were almost exclusively undergraduate students at the Louisiana State University who received extra credit toward their grade in psychology courses. Only subjects with complete item data (N = 671) on the 20 masculinity and 20 femininity items of the BSRI were used in the initial factor analyses; the structural regression models (see below) analyzed data from the 671 persons with complete data on a subset of BSRI items.

The cross-sectional sample of 671 adults (388 females, 283 males) ranged in age from 18 to 91 (M = 39.34, SD = 16.86). The sample tended to be relatively well educated for these birth cohorts (mean years of education = 14.44, SD = 2.41), with slightly fewer years of education in the oldest age decades. For model development purposes, the participants were stratified on gender and then randomly assigned to two half-samples. The first half-sample (hereafter, the exploratory sample; N = 336) was used to develop an appropriate item factor model for the BSRI; the second half-sample (validation sample; N = 335) was used to cross-validate the final model from the exploratory sample.

Materials

The BSRI contains 60 descriptive adjectives that individuals rate on a 7-point Likert-scale (1: *Never or almost never true*; 7: *Always to almost always true*). Respondents were instructed: "Below is a list of words that could be used to describe an individual. Please indicate in the space next to each word the degree to which you believe that word describes you."

According to Bem (1974), 20 BSRI items measure masculinity, 20 items measure femininity, and the remaining 20 items are gender-neutral fillers. Table I provides a list of the adjectives, their scale assignments, and descriptive statistics from the present sample (means, standard deviations, and correlations with gender and age).

Table I. BSRI Items, Means, and Standard Deviations (SD) for Total Sample, Correlation with Gender and Age

Item	BSRI Scale[a]	Mean	*SD*	Item correlation	
				With gender	With age
1. Self-reliant	M	5.70	1.11	−.20[c]	.12[c]
2. Yielding	F	4.57	1.14	.03	.07
3. Helpful	N	5.85	.90	.08[b]	.19[c]
4. Defends own beliefs	M	5.73	1.12	−.03	.11[c]
5. Cheerful	F	5.40	1.02	.05	.08[b]
6. Moody	N	3.70	1.34	.05	−.24[c]
7. Independent	M	5.71	1.15	−.08[b]	.06
8. Shy	F	3.45	1.56	.02	−.15[c]
9. Conscientious	N	5.74	1.15	.11[c]	.24[c]
10. Athletic	M	3.96	1.87	−.34[c]	−.28[c]
11. Affectionate	F	5.56	1.16	.15[c]	−.01
12. Theatrical	N	3.05	1.62	−.02	−.11[c]
13. Assertive	M	4.80	1.34	−.09[b]	.00
14. Flatterable	F	4.24	1.36	.01	−.20[c]
15. Happy	N	5.51	.98	.01	.11[c]
16. Strong personality	M	5.27	1.25	−.08[b]	.03
17. Loyal	F	6.25	.92	.06	.20[c]
18. Unpredictable	N	3.63	1.54	.04	−.19[c]
19. Forceful	M	4.14	1.38	−.13[c]	.12[c]
20. Feminine	F	3.90	2.32	.89[c]	.00
21. Reliable	N	6.17	.91	.10[b]	.25[c]
22. Analytical	M	4.88	1.46	−.16[c]	.01
23. Sympathetic	F	5.57	1.17	.21[c]	.22[c]
24. Jealous	N	3.50	1.52	−.04	−.27[c]
25. Leadership abilities	M	5.11	1.38	−.20[c]	−.08[b]
26. Sensitive to others	F	5.62	1.04	.17[c]	.17[c]
27. Truthful	N	6.15	.87	.05	.25[c]
28. Willing to take risks	M	4.72	1.38	−.18[c]	−.10[c]
29. Understanding	F	5.65	.97	.15[c]	.06
30. Secretive	N	3.74	1.58	.03	−.13[c]
31. Makes decisions easily	M	4.68	1.31	−.21[c]	.08[b]
32. Compassionate	F	5.56	1.03	.21[c]	.17[c]
33. Sincere	N	5.95	.95	.08[b]	.23[c]
34. Self-sufficient	M	5.61	1.18	−.11[c]	.11[c]
35. Eager to soothe	F	5.45	1.23	.16[c]	.09[b]
36. Conceited	N	2.71	1.40	−.13[c]	−.17[c]
37. Dominant	M	3.99	1.56	−.18[c]	−.11[c]
38. Soft-spoken	F	4.08	1.56	−.01	.16[c]
39. Likable	N	5.36	.98	.01	−.04

Table I. Continued

Item	BSRI Scale[a]	Mean	*SD*	Item correlation	
				With gender	With age
40. Masculine	M	3.48	2.32	−.87[c]	.00
41. Warm	F	5.38	1.02	.11[c]	.09[b]
42. Solemn	N	3.98	1.31	−.10[b]	−.02
43. Willing to take stand	M	5.28	1.16	−.14[c]	.08
44. Tender	F	5.22	1.18	.18[c]	.08[b]
45. Friendly	N	5.89	.96	.04	.03
46. Aggressive	M	4.46	1.44	−.18[c]	.07
47. Gullible	F	3.38	1.49	.20[c]	−.13[c]
48. Inefficient	N	2.63	1.23	.04	.00
49. Acts as a leader	M	4.76	1.43	−.24[c]	−.03
50. Childlike	F	2.82	1.38	.03	−.31[c]
51. Adaptable	N	5.18	1.15	−.05	−.04
52. Individualistic	M	5.13	1.26	−.06	−.09[b]
53. No harsh language	F	4.37	1.77	.04	.11[c]
54. Unsystematic	N	3.10	1.42	.10[b]	−.02
55. Competitive	M	4.72	1.47	−.31[c]	−.16[c]
56. Loves children	F	6.00	1.22	.11[c]	.17[c]
57. Tactful	N	5.09	1.19	.07	.09[b]
58. Ambitious	M	5.39	1.23	−.11[b]	−.12[c]
59. Gentle	F	5.42	1.11	.13[c]	.09[b]
60. Conventional	N	4.81	1.26	.07	.13[c]

[a]BSRI scale assignment from Bem (1974) (M = Masculinity, F = Femininity, N = Neutral; females were coded as 2, males as 1).
[b]$p < .05$.
[c]$p < .01$.

As our purpose was to analyze the factor structure of the masculinity and femininity items from the BSRI, the gender-neutral items were omitted from further analysis. In addition, 4 adjectives ("flatterable"—Item 14; "gullible"—Item 47; "childlike"—Item 50; and "does not use harsh language"—Item 53) normally assigned to the BSRI Femininity scale were excluded from the analysis because they did not correlate significantly with other BSRI items (see also Pedhazur & Tetenbaum, 1979). Thus the factor analyses focused on the remaining 36-item subset of the BSRI.

Procedure

We used the LISREL VI program (Jöreskog & Sorbom, 1984) to conduct the restricted common factor analyses and structural equation models reported in this paper (see Hayduk, 1987, for a detailed description of the LISREL model and its use). The initial item factor analyses were specified using the standard LISREL measurement model for the item covariance matrix *S*,

$$S = LX \cdot PH \cdot LX' + TD \tag{1}$$

where *LX* is a factor pattern matrix of (unstandardized) regressions of variables on factors, *PH* is a factor covariance matrix, and *TD* is a residual covariance matrix. Given that metric solutions are difficult to interpret, parameter estimates reported in the tables have been rescaled (standardized) to assist in interpretation. Simultaneous analyses in the two gender groups were also specified as models fitting the item covariance matrix. These multiple-group analyses' parameter estimates have been rescaled using Jöreskog's (1971) quasi-standardized metric.

Hierarchical factor models and structural equation models were fitted on LISREL's *y* side, using parameter matrices *BE, LY, PS,* and *TE,* with second-order loadings and/or structural regression coefficients modeled in the BE matrix (see Hayduk, 1987). In LISREL notation, the covariance matrix *S* is now structured as

$$S = LY \cdot [(I - BE)^{-1} \cdot PS \cdot (I - BE)^{-1\prime}] \cdot LY' + TE \tag{2}$$

where *LY* is the factor pattern matrix, *TE* is the residual covariance matrix (for observed variables), *BE* is the structural regression matrix, containing directed relations between latent variables, and PS is the covariance matrix of regression residuals. Equation (2) can be understood as nesting the structural model inside the equation for the measurement model (see McDonald, 1978). Equation (2) shows that any structural model can be treated as being a more restricted variant of the measurement model based upon the same factor analysis specification; the factor covariance matrix of Eq. (1) *PH* has been structured as a function of regression coefficients in *BE* and the covariance matrix of regression residuals *PS*.

The model development approach used here is best characterized as a specification search, using a covariance structures fitting algorithm to conduct exploratory research (see Hertzog, 1990). We evaluated models based upon goodness-of-fit statistics, patterns of normalized residuals (scaled discrepancies between sample covariances and fitted covariances from the estimated LISREL model), LISREL modification indices, interpretability of parameter estimates, and theoretical considerations. In this kind of specification search, the likelihood ratio (LR) χ^2 is best understood as an index of fit rather than as a test of a specific hypothesis. In some cases differences in LR χ^2 between nested models may be treated as tests of specific statistical hypotheses. In the present paper the most important use of such tests is for the simultaneous analysis of the equivalence of males' and females' item factor structure (e.g., tests of the equality of factor loadings between male and female groups). Such differences in χ^2 in the multiple groups analyses should be treated as tests of the omnibus null hypothesis that an entire matrix of parameters is equivalent, and as such one must still inquire about the source of the significant differences. We employ specific post hoc tests on individual parameters in order to identify the specific parameters that differ between males and females. Given that we are comparing parameters across independent samples, specific comparisons may be made using the following z test:

$$z = (PE_1 - PE_2) / (SE_1 2 + SE_2 2)^{.5} \quad (3)$$

where PE denotes the LISREL parameter estimate, SE denotes the corresponding standard error of estimate, and the subscript denotes group membership.

The LR χ^2 is known to be sensitive to sample size, and additional indices of fit are required. We report four different fit indices, each with their own advantages and disadvantages (see Bollen & Long, 1993): (a) LISREL's goodness-of-fit index (GFI; Jöreskog & Sorbom, 1984); (b) the relative noncentrality index (RNI; McDonald & Marsh, 1990), which can be treated as an unbiased estimate of the proportion of information in the covariances accounted for by the model; (c) the Tucker-Lewis index (TLI; McDonald & Marsh, 1990); and (d) the single-sample cross-validation index (ECVI; Browne & Cudeck, 1989), which is essentially a rescaled Akaike Information Criterion. The ECVI imposes a penalty on the fit index for the number of free parameters estimated.

When structural equation models or hierarchical factor models are analyzed, the fit indices just listed reflect the fit to the moment matrix being analyzed—in this case, the BSRI item covariance matrix. It is also useful to determine the extent to which the structural model fits the first-order factor covariance matrix [as would be estimated in a standard confirmatory factor analysis, namely, the PH matrix of Eq. (1)]. We therefore report a relative goodness-of-fit index (RFI; Mulaik, James, Van Alstine, Bennett, Lind, & Stillwell, 1989), analogous to the Bentler and Bonett (1980) normed fit index, for models with structural regressions on the first-order item factors. The RFI may be interpreted as the proportion of information in the first-order factor covariances accounted for by the structural model. Although there are no hard and fast objective rules for adequate model fit, a commonly accepted rule of thumb for all the fit indices except the ECVI is that a fit index of .9 or greater represents a fit sufficiently adequate to consider retaining a model. Browne and Cudeck (1989) recommend selecting the model with the lowest ECVI. Model adequacy cannot, however, be judged solely on the basis of overall fit, but must also be judged on factors such as salience and interpretability of parameter estimates (Hayduk, 1987).

Our general attack on the problem was to search for a model in the exploratory half-sample, and then cross-validate that model in the validation half-sample. The final model was then re-estimated in the total sample. We opted to conduct the comparative factor analysis in males and females using the entire sample, segregated by gender. This approach provides more statistical power for detecting differences between the groups. Cross-validation would have been impractical, given the need to maintain maximum sample size in each gender group (and considering the large number of parameters estimated in a factor analysis of 36 items). Structural equation models were also estimated in the total sample, for similar reasons.

RESULTS

Masculinity and Femininity Self-Ratings

As can be observed in Table I, self-ratings of the Masculine and Feminine items correlated highly with gender. As would be expected, males rated themselves higher on masculinity and females rated themselves higher on femininity. Figure 1 displays the frequency polygon for responses on these two items for the male and female samples. The gender-consistent response patterns generated a bimodal distribution of masculinity and femininity self-ratings in the aggregated sample, helping to produce the high correlations with gender reported in Table I. These distributions suggest a possible cause for the high association of masculinity and femininity ratings in gender-aggregated samples. In the present data, the aggregated correlation of Masculinity with Femininity was –.87. Correlations of masculinity and femininity ratings were considerably lower when computed within each gender group ($r = -.49$ in males, $r = -.42$ in females). These results support Pedhazur and Tetenbaum's (1979) argument that BSRI items should be factored separately in male and female groups, especially when the Masculinity and Femininity items are included. In order to compare results with other studies, however, we opted to retain these items in the initial models run on the gender-aggregated half-samples.

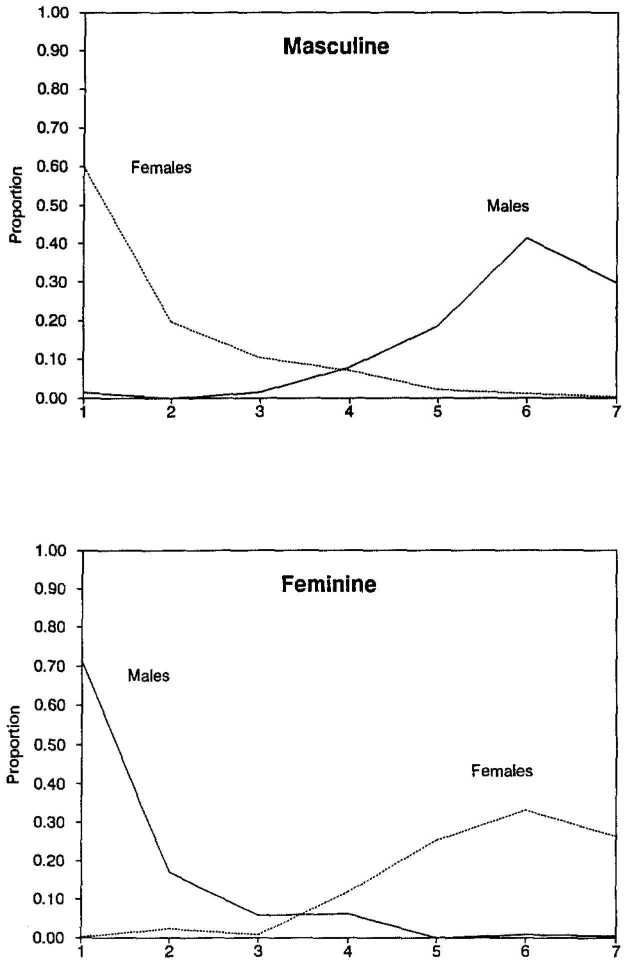

Fig. 1. Distribution of males' (N = 277) and females' (N = 382) self-ratings on BSRI items Masculine and Feminine (1: *Never or almost never true*; 7: *Always or almost always true*).

Initial CFA Models

The first set of analyses used the exploratory half-sample to develop a model for the 36 BSRI items. As expected, a two-factor model specifying a Masculinity factor and a Femininity factor fit the data poorly (χ^2 = 2229.98, df = 593, GFI = .729, RNI = .693, TLI = .674, ECVI = 7.13).

A four-factor model, based on an a priori assignment of items to factors conceptualized as Instrumental, Interpersonal Sensitivity, Self-Sufficient, and bipolar Masculinity/Femininity fit better than the two-factor model (χ^2 = 1462.37, *df* = 583, GFI = .798, RNI = .835, TLI = .823, ECVI = 4.49). LISREL's modification indices suggested the existence of two large residual relationships between two pairs of items, 4 (Defends own beliefs) with 43 (Willing to take a stand), and 25 (Has leadership abilities) with 49 (Acts as a leader). In both cases, item-specific associations due to similar wording or concepts seemed plausible. A model estimating residual covariances between these pairs of items improved the fit but still suggested that additional factors would be required to fit the item covariance structures adequately.

At this point we proceeded to develop a model by testing a number of alternative models based upon inspection of fit to the sample data, as indicated by residuals and LISREL modification indices. For example, we identified two subsets of large normalized residuals among the items loading on the Interpersonal Sensitivity factor. A model splitting the majority of Femininity items into two factors resulted in improvement in the fit of the model to the data. Tentatively, a nine-factor specification was adopted as the best fitting model for the exploratory sample. It was then tested in the confirmatory half-sample. A principal feature of this model was that three items were modeled as single indicator "factors"—Masculine, Feminine, and Analytic. The fit of the model had been improved by separating the first 2 items, despite their high (negative) correlation with each other, instead of specifying a bipolar Masculinity/Femininity factor. The nine-factor model also specified 3 femininity items (Yielding, Shy, Soft-spoken) to load (negatively) on a masculinity factor identified as Dominant (marked by loadings for items such as Aggressive and Dominant). This model did not fare as well in cross-validation, however, and a ten-factor model in which these 3 femininity items loaded on a separate factor of their own was finally selected for further analysis.

The final ten-factor model fit reasonably well in the exploratory sample (χ^2 = 983.69, *df* = 544, GFI = .859, RNI = .918, TLI = .905, ECVI = 3.68), the confirmatory sample (χ^2 = 983.07, *df* = 544, GFI = .854, RNI = .917, TLI = .904, ECVI = 3.72), and the total sample (χ^2 = 1209.01, *df* = 544, GFI = .906, RNI = .936, TLI = .926, ECVI = 2.19). We also analyzed the data separately in male and female subsamples. Based upon goodness-of-fit indices, the model fit the data for the subsample of 386 women slightly better than the subsample of 281 men (χ^2 = 961.11, *df* = 544, GFI = .881, RNI = .960, TLI = .953, ECVI = 3.13; vs. χ^2 = 952.62, *df* = 544, GFI = .841, RNI = .961, TLI = .954, ECVI = 4.27, respectively), but the fit was deemed adequate for both gender groups.

The first three columns of Table II report the standardized factor loadings for the seven multiple-item factors for the aggregated sample and for the male and female subsamples. Four of the factors (Decisive, Dominant, Athletic, and Self-Sufficient) involved items from the BSRI masculinity item set, and the other three factors (Interpersonal Affect, Compassionate, and Shy) were defined by BSRI femininity items. As can be seen in Table II, these standardized factor loadings were obviously quite similar in the male and female subsamples.

Table II. Standardized Factor Loadings for Aggregated, Female, and Male Samples

	Samples			
Factor and item	Aggregated	Female	Male	Simultaneous
Interpersonal affect				
Yielding	.27	.20	.41	.29
Cheerful	.51	.49	.54	.51
Affectionate	.67	.63	.72	.66
Loyal	.49	.53	.44	.50
Soft-spoken	.49	.32	.50	.37
Warm	.80	.80	.80	.80
Tender	.77	.81	.71	.76
Loves children	.46	.44	.47	.45
Gentle	.78	.77	.85	.79
Decisive				
Defends own beliefs	.48	.47	.52	.49
Assertive	.43	.43	.25	.36
Strong personality	.68	.68	.69	.68
Has leadership ability	.70	.68	.69	.68
Willing to take risks	.48	.45	.44	.45
Makes decisions easily	.56	.52	.60	.53
Willing to take a stand	.65	.62	.68	.64
Aggressive	.34	.32	.42	.35
Acts as leader	.71	.70	.68	.69
Ambitious	.55	.53	.60	.55
Shy				
Yielding	.27	.27	.29	.26
Shy	.56	.61	.51	.55
Soft-spoken	.62	.55	.76	.63
Gentle	.25	.17	.36	.25
Self-sufficient				
Self-reliant	.71	.72	.69	.70
Independent	.73	.77	.66	.74
Self-sufficient	.75	.77	.72	.75
Individualistic	.52	.54	.46	.52

Table II. Continued

Factor and item	Samples			
	Aggregated	Female	Male	Simultaneous
Athletic				
Athletic	.77	.84	.74	.90
Competitive	.48	.34	.53	.35
Dominant				
Assertive	.32	.30	.54	.39
Forceful	.70	.67	.71	.69
Dominant	.69	.67	.69	.67
Aggressive	.44	.46	.34	.42
Competitive	.42	.50	.34	.48
Compassionate				
Sympathetic	.73	.71	.72	.71
Sensitive to needs	.74	.68	.80	.73
Understanding	.72	.75	.68	.72
Compassionate	.80	.77	.82	.79
Eager to soothe	.68	.65	.69	.67

Table III reports the factor correlations for the total sample. There are several noteworthy patterns in the correlations. First, Decisive and Dominant were highly correlated, as were IAF and Compassionate, but for both pairs of factors the exploratory models detected significant improvement of fit when the factors were differentiated. Second, the major femininity factors, IAF and Compassionate, correlated positively with the major masculinity factors, Decisive and Dominant. This pattern is inconsistent with the hypothesis that factors from BSRI's masculinity and femininity item pools are necessarily negatively correlated, and perhaps even polar opposites. However, Shy did correlate negatively with the masculinity factors. Finally, the Self-Sufficient and Analytic factors were successfully differentiated from the Decisive and Dominant factors, although there were sizable positive correlations between these four factors. Factor correlations were similar in pattern for males and females, but salient differences existed that were analyzed in more detail as part of the multiple group analyses reported next.

Comparative Analyses of Male and Female Subsamples

In order to test for gender differences in BSRI item factor structure, a set of simultaneous factor analyses were conducted on the two gender

Table III. Factor Correlations for Ten-Factor Model in the Combined Sample[a]

Factors	2	3	4	5	6	7	8	9	10
1. M	–.18[c]	.31[c]	–.03	.17[c]	.52[c]	.21[c]	.33[c]	–.22[c]	–.87[c]
2. IAF	—	.36[c]	–.13[c]	.25[c]	.02	.06	–.08[b]	.86[c]	.30[c]
3. DEC		—	–.64[c]	.72[c]	.47[c]	.35[c]	.85[c]	.26[c]	–.21[c]
4. SHY			—	–.20[c]	–.11[c]	.05	–.54[c]	–.01	.00
5. S-S				—	.23[c]	.36[c]	.50[c]	.22[c]	–.10[c]
6. ATH					—	.19[c]	.25[c]	–.10[c]	–.45[c]
7. ANC						—	.34[c]	.09[b]	–.15[c]
8. DOM							—	–.09[b]	–.23[c]
9. COM								—	.31[c]
10. F									—

[a]M: Masculine; IAF: Interpersonal Affect; DEC: Decisive; S-S: Self-Sufficient; ATH: Athletic; ANC: Analytic; DOM: Dominant; COM: Compassionate; F: Feminine.
[b]$p < .05$.
[c]$p < .01$.

groups. The first simultaneous model imposed equality constraints on the factor pattern matrices, forcing the factor loadings to be equal for males and females. This model fit the data relatively well ($\chi^2 = 1963.63$, $df = 1120$, $GFI_{females} = .879$, $GFI_{males} = .837$, RNI = .910, TLI = .899). To test the hypothesis of equal factor loadings, a baseline fit measure was constructed by summing the χ^2 for the two models estimating the base ten-factor model separately for the male and female subsamples. The difference in χ^2 between this baseline and the model imposing equality constraints was small (change in $\chi^2 = 39.90$, $df = 32$, $p > .10$), indicating that the hypothesis of equivalent factor loadings could not be rejected.

The right-hand column of Table II reports the rescaled factor loadings, using a common metric rescaling (Jöreskog & Sorbom, 1988) to preserve gender equivalence in the loadings. These rescaled loadings were highly similar to the standardized loadings obtained from analysis of the total gender-aggregated sample.

Given gender equivalence in factor loadings, the next step was to determine whether there were gender differences in the relationships between the BSRI item factors. Thus, a second model imposed equality constraints on both the factor covariance matrices and factor pattern matrices. This model did not fit as well ($\chi^2 = 2169.00$, $df = 1175$, $GFI_{females} = .869$, $GFI_{males} = .820$, RNI = .895, TLI = .888), and the difference in fit was statistically significant ($\chi^2 = 205.37$, $df = 55$, $p < .001$). The hypothesis of equal factor covariance matrices between the two groups was therefore rejected.

In order to better localize the differences, we partitioned the factor covariance matrices into two parts: (a) variances and covariances associated with the Masculine and Feminine items and (b) variances and covariances associated with the remaining eight gender-related attribute factors. The first set of models tested gender differences in the parameters involving the Masculine and Feminine items. A model relaxing constraints on these parameters fit significantly better than a model constraining the entire factor covariance matrix to be equal (χ^2 = 124.70, *df* = 19, $p < .001$), suggesting that a substantial portion of the gender differences in factor variances and covariances was associated with the Masculine and Feminine items.

Differences in factor covariance matrices could reflect (a) gender differences in variability of factor scores, as reflected in factor variances, (b) gender differences in factor correlations, or (c) both. To address this issue, we specified a model that directly estimated standardized factor correlations and factor standard deviations instead of unconstrained factor covariance matrices (see Marsh & Hocevar, 1985). A model forcing gender equivalence of the Masculine and Feminine item factors' correlations with the other eight factors, while leaving the factor standard deviations unconstrained across groups, fit significantly worse than the model with no constraints on the item factor covariance matrix (χ^2 = 2059.32, *df* = 1137, GFI_{males} = .829, $GFI_{females}$ = .874, RNI = .902, TLI = .892). The difference between these two models tested the null hypothesis of equal factor correlations for the two gender groups, which was rejected (change in χ^2 = 95.69, *df* = 17, $p < .001$). Males and females differed in the magnitude of correlations between the Masculine and Feminine items and the other BSRI item factors.

The pattern of correlations differed markedly between males and females. As can be seen in Table IV, a large number of the *z* tests for gender differences in the correlations were statistically significant. For males, self-rated masculinity is associated positively with scores on Decisive and Dominant, as expected. However, males' self-rated masculinity ratings also correlated positively with IAF and Compassionate. Females' self-rated masculinity correlated negatively with these factors (hence the significant gender differences in these correlations). However, females self-rated femininity correlated more highly with IAF and Compassionate than did males' self-rated femininity. In general, then, gender-consistent ratings (i.e., males' ratings of Masculine, females' ratings of Feminine) correlated higher with self-ratings on positive gender-related attributes for each gender group. However, males who view themselves as masculine are more likely to rate themselves as interpersonally sensitive than women who rate themselves as masculine.

Table IV. Correlations of BSRI Masculine and Feminine Items with Other Item Factors for Female and Male Subsamples[a]

	Masculine self-rating			Feminine self-rating		
	Females	Males		Females	Males	
Factor	(SE)	(SE)	z	(SE)	(SE)	z
IAF	−.29(.05)	.25(.06)	−6.81[c]	.53(.04)	−.01(.07)	7.11[c]
DEC	.06(.06)	.40(.06)	−4.26[c]	.12(.06)	−.20(.06)	3.83[c]
SHY	.01(.07)	−.36(.07)	3.69[c]	.07(.07)	.13(.08)	−.58
S-S	.04(.06)	.25(.07)	−2.44[b]	.11(.06)	−.09(.07)	2.22[b]
ATH	.32(.07)	.17(.06)	1.67	−.16(.06)	−.03(.06)	−1.47
ANC	.12(.05)	.19(.06)	−1.01	.01(.05)	−.03(.06)	.44
DOM	.25(.06)	.39(.06)	−1.68	−.12(.06)	−.13(.07)	.13
COM	−.20(.05)	.20(.06)	−4.99[c]	.36(.05)	−.05(.06)	5.11[c]

[a]See footnote *a* of Table III.
[b]$p < .05$.
[c]$p < .01$.

Next, our attention focused on the factor variances and covariances among the remaining 8 BSRI item factors. The appropriate model comparison rejected the null hypothesis of gender equivalence for these parameters ($\chi^2 = 80.67$, $df = 36$, $p < .01$). To further examine these differences, we estimated another model, directly estimating factor correlations (as in the previous set of models), and imposing gender equivalence on all factor correlations. This model fit significantly worse than the model with constraints on only the correlations involving the Masculine and Feminine items, rejecting the null hypothesis of gender equivalence in the remaining BSRI item factor correlations ($\chi^2 = 2131.41$, $df = 1165$, $GFI_{females} = .870$, $GFI_{males} = .823$, RNI = .897, TLI = .889; change in $\chi^2 = 72.09$, $df = 28$, $p < .01$).

Follow-up z tests suggested that three pairs of factor correlations differed across the gender groups: Dominant with Decisive (.91 for females, .78 for males, $z = 2.03$, $p < .05$), Dominant with Athletic (.06 for females, .27 for males, $z = -1.98$, $p < .05$), and Compassionate with Self-Sufficient (.35 for females, .13 for males, $z = 2.38$, $p < .05$). Gender differences in the correlations of Dominant with Self-Sufficient just failed to achieve statistical significance (.55 for females, .39 for males, $z = 1.86$, $p < .10$). The remaining correlations were similar in magnitude and approximated the estimates in the gender-aggregated sample (see Table III). These differences were, therefore, not as striking as the differences in factor correlations involving Masculine and Feminine items.

The finding of gender equivalence in BSRI item factor loadings provides evidence that the same factors are present in both males and females, and that these factors have equivalent relationships to BSRI items. The factors appear to differ somewhat in their interrelationships, especially with the Masculine and Feminine self-ratings. However, the factors can be considered to be qualitatively invariant in terms of their item definitions. With respect to the gender differences in factor covariances, the localization of gender differences to the ratings of masculinity and femininity suggested that it was appropriate to conduct the hierarchical and regression models reported below using the total gender-aggregated sample, especially if the Masculine and Feminine items were removed from the analysis. Indeed, the limited differences in factor covariances not involving the masculinity and femininity ratings could easily be attributed to selection processes and their effects on factor covariances, given that males and females differ significantly in BSRI item means (see Meredith, 1964, and Mulaik, 1972, for further explanations).

Hierarchical Models for BSRI Item Factors

One obvious aspect of the item factor analyses was the high correlations between several item factors, most notably Decisive with Dominant and Interpersonal Affect with Compassionate. Factor correlations greater than .8 would in many cases be considered grounds for combining factors, although these factors were empirically differentiable, in terms of producing salient improvement in fit of the item factor models.

Other models for the BSRI contain far fewer factors. Marsh and Myers (1986) postulate two common factors—global Masculinity and global Femininity. Pedhazur and Tetenbaum (1979) identified four factors, including the bipolar MF rating factor. A satisfactory resolution of these apparent differences involves the hypothesis that the BSRI item factors extracted here have a higher order factor structure. In order to address this issue, we proceeded to conduct simultaneous hierarchical factor analyses, which modeled the BSRI item factor covariances as being determined by higher order factors. Our chief interest here was to test a hierarchical analog of the two-factor (Masculinity, Femininity) solution favored by Marsh and Myers (1986) against a hierarchical model differentiating subfactors among the masculinity items, as in the exploratory analysis of Pedhazur and Tetenbaum (1979).

Simultaneous hierarchical factor models have been successfully employed in analyses of other self-report questionnaires (Hertzog, Van Alstine, Usala, Hultsch, & Dixon, 1990; Marsh, 1985; Marsh & Hocevar, 1985; Tanaka & Huba, 1984). They are in principle similar to first-order factor analyses of item subscales, but the first-order item factor covariances that are implicitly factored to estimate second-order loadings are disattenuated for measurement error. Moreover, the hierarchical model is methodologically preferable in the present case for two reasons, both related to problems of aggregation bias in item subscales: (a) that specific residual covariances among items were detected in the item factor analysis, and (b) that a few items loaded on more than one factor. Aggregating item responses into unit-weighted item subscales would embed these effects in subscales with no adequate means to account for these types of influences. Conversely, the hierarchical model estimates second-order factor loadings while allowing these specific relationships to remain as part of the more complex first-order factor specification.

For the purpose of estimating the hierarchical solution, the Masculinity, Femininity, and Athletic items were deleted from the analysis, leaving 33 BSRI items to be factored.[3] A total sample 668 persons had complete data for these 33 items. We started with a first-order factor analysis of these items, using a specification for the reduced item set that was otherwise identical to the final model for the combined sample. We fit seven factors (IAF, Decisive, Shy, Self-Sufficient, Dominant, Compassionate, and Analytic [marked by the single BSRI item, Analytical]). The model fit the data relatively well (Model H1, Table V). It provided a baseline against which to evaluate any hierarchical model of interest. Because it freely estimated all first-order factor covariances, it (trivially) accounted for all the information contained in these covariances. A hierarchical model, on the other hand, places restrictions on these covariances as part of its second-order factor specification. Hence any restricted hierarchical model could fit no better than H1. It could therefore be used as a baseline for evaluating the adequacy of the second-order factor specification in any hierarchical model of interest (see Rindskopf & Rose, 1988). We also estimated a null model (Model H2) in the first-order factor covariances by estimating orthogonal first-order factors (i.e., by fixing all factor covariances to equal zero). This model fit poorly (see Table V). Models H1 and H2 define a range of best and worst possible fits for the hierarchical models. They were used to estimate the RFI index for the proportion of information in the first-order factor covariances accounted for by the hierarchical models (Mulaik et al., 1989).

The next hierarchical model specified two higher order factors, Masculinity and Femininity, basically in accord with patterns predicted by BSRI

[3]The M and F items were deleted because they appear to generate different correlations in males and females, and because of their bimodal distribution in a gender-aggregated sample. The Athletic item was eliminated because the Athletic factor had small correlations with other BSRI item factors. It would have contributed little to the definition of the second-order factor structure.

Table V. Goodness-of-Fit Statistics: Hierarchical Factor Models

Model	χ^2	*df*	GFI	RNI	TLI	ECVI	RFI[a]
H0 Null model	9507.70	528	.304	—	—	—	—
H1 Seven-factor model	1114.91	468	.906	.928	.919	1.95	—
H2 Orthogonal factors	2759.19	489	.804	.747	.727	4.35	—
H3 M and F factors	1293.79	481	.893	.909	.901	2.18	.891
H4 M and F; Dominant on F	1201.24	480	.899	.920	.912	2.04	.947
H5 Three second-order factors	1138.83	476	.903	.926	.918	1.96	.985
H6 H4 with fixed variances	1223.22	482	.898	.917	.910	2.07	.934

[a]For fit to first-order factor covariances.

item assignments to these scales (e.g., Bem, 1974). Masculinity determined the following first-order factors: Decisive, Dominant, Self-Sufficient, Shy, and Analytic. Femininity determined IAF and Compassionate. This model, denoted H3, fit somewhat worse than Model H1 (see Table V), although it did account for a relatively good proportion of covariance in BSRI item factors (RFI = .891). Estimated residual variances for Decisive and IAF were negative. All estimated factor loadings were statistically significant, and the estimated covariance among Masculinity and Femininity was significant and positive (a rescaled factor correlation of .31). There was a large modification index for the loading of Dominant on Femininity (74.58)—suggesting that the relationship of Dominant to Femininity was less positive than implied by the estimated covariance between the two higher order factors. A new model, H4, adding the loading of Dominant on Femininity, fit significantly better than H3, and the new second-order loading was negative and significant. The negative residual variances for Decisive and IAF remained. All the fit indices showed improvement from the original two-factor model, especially the RFI. There was still a salient difference between Models H4 and H1 ($\chi^2 = 86.33$, $df = 12$, $p < .001$), but the RFI indicated that about 95% of the information in the latent covariances was accounted for by H4.

The next model, H5, specified a three-factor model. Analytic, Self-Sufficient, and Shy loaded on the new third factor, although the latter two factors still marked the Masculine factor as well. This model fit slightly better than H4, even when considering the Tucker-Lewis and ECVI fit indices that adjust for the number of parameters estimated (see Table V).

However, the residual variance for Analytic was also negative. Both freely estimated factor loadings were significantly different from zero. Furthermore, Self-Sufficient had a somewhat higher standardized loading on Masculinity than on the new third factor. Although it appeared that there might be additional relationships between BSRI item factors not fully accounted for by the two-factor solution, it also appeared that the source of these additional relationships was underdetermined by the BSRI item factors. We therefore opted to select H4 as the preferred model.[4] A final model, H6, fixed the negative residual variances for Decisive and IAF to zero.

Table VI reports the unstandardized factor loadings, their standard errors, and the rescaled (standardized) loadings from Model H6. All the second-order factor loadings were statistically significant. First-order loadings were quite comparable to the solution for the item factor analysis on the combined sample, as well as for Model H1, and hence are not reported. The Masculinity and Femininity factors correlated .35 with each other.

Table VI. Second-Order Factor Solution for BSRI Item Factors[a]

	Factor pattern matrix					
	IAS		INS		ANL	
Factors	(SE)	[b]	(SE)	[b]	(SE)	[b]
IAF	1.00(—)[c]	1.00	0(—)[c]	0	0(—)[c]	0
DEC	0(—)[c]	0	1.00(—)[c]	1.00	0(—)[c]	0
SHY	0(—)[c]	0	−1.28(.17)	−1.02	.68(.14)	.69
S-S	0(—)[c]	0	.53(.08)	.46	.33(.07)	.37
ANC	0(—)[c]	0	0(—)[c]	0	1.00(—)[c]	.70
DOM	−.57(.06)	−.42	1.56(.13)	1.01	0(—)[c]	0
COM	.79(.05)	.86	0(—)[c]	0	0(—)[c]	0

[a]IAS: Interpersonal Sensitivity; INS: Instrumental; ANL: Analytical. For other abbreviations, see footnote [a] of Table III.
[b]Standardized regression coefficient.
[c]Denotes fixed parameter, with no standard error of estimate.

[4]There were substantial modification indices for residual covariances between the Analytic and Shy item factors in H4 (which was one indication that the three-factor model might have been viable). We also tried an alternative model that fit a residual covariance of Self-Sufficient and Analytic (the largest modification index). This model fit better than H4 (χ^2 = 1189.96, df = 481, GFI = .901) and the estimated residual covariance was statistically significant. However, this alternative specification led to nonconvergent solutions when used in a preliminary structural regression model (see next section). Under the circumstances we concluded that the fit of the revised two-factor model was sufficiently good to use it for structural regression, even though there were some indications of possible additional relationships among the BSRI item factors it does not fully account for.

Structural Regression Model with Age and Gender

The next set of analyses were designed to estimate the relationship of age and gender with the BSRI item factors by adding these two variables to the hierarchical factor model. Given the wide variability of age range, we were interested in whether one could model age and gender differences in BSRI items as being mediated by the BSRI item factors. That is, are there specific age and gender differences in BSRI items above and beyond age and gender differences on the first-order and second, order BSRI factors?

We began by adding age and gender as factors to the seven-factor measurement model for the hierarchical solution (Model H1 in the preceding section). This model specified free covariances of age and gender with the seven BSRI first-order factors, but allowed no additional BSRI item relationships to age and gender. This model, R1, had a relatively good fit (see Table VII), but residuals and modification indices suggested that several items had specific relationships to age and gender that could not be modeled as being mediated by covariances between these variables and the first-order BSRI item factors. The model did account for much of the gender and age differences in BSRI items, but not all of the gender and age-related variance in the BSRI could be accounted for by the BSRI common factors. We then computed a series of models adding specific regressions of BSRI items on age and gender in addition to the regressions of the BSRI item factors on age and gender. These additional regression parameters are best understood as reflecting residual relationships that are independent of the gender and age-related variance that can be accounted for by the BSRI item factors. The final model in this series, Model

Table VII. Goodness-of-Fit Statistics: Structural Regression Models

Model	χ^2	*df*	GFI	RNI	TLI	ECVI	RFI[a]
R0 Null model	9916.06	595	.310	—	—	—	—
R1 Seven-factor model with age and gender	1373.32	520	.889	.908	.895	2.39	—
R2 Specific age and gender regression	1195.10	513	.905	.927	.915	2.14	—
R3 R2 with orthogonal factors	2979.35	549	.798	.739	.717	5.17	—
R4 Age and gender predict M and F	1389.79	537	.892	.909	.899	2.36	.891
R5 Final model	1341.73	534	.895	.913	.903	2.30	.918

[a]For fit to first-order factor covariances.

R2, specified regressions of BSRI Items 19, 52, 55, 58, 17, and 38 on age, and BSRI Item 55 on gender. Table VIII reports the rescaled factor correlations with age and gender for these two models (with and without specific item relationships). Although these values did change slightly between models, the general pattern is consistent. Age yielded small but significantly positive relationships with IAF, Self-Sufficient, and Compassionate. Conversely, age yielded a negative relationship to Shy, with older persons reporting lower levels of shyness. The age patterns of specific item relationships were generally consistent with the factor correlations; age was associated with higher endorsement of BSRI Items 17 (Loyalty), 19 (Forceful), and 38 (Soft-Spoken), and with lower levels of BSRI Items 52 (Individualistic), 55 (Competitive), and 58 (Ambitious). Gender, as expected, had significant relationships with all BSRI item factors (except Shy). It should be noted, however, that the magnitude of these correlations is relatively small; all of the rescaled correlations with BSRI factors were lower than .3 in absolute value. The correlations with age were likewise rather small in magnitude.

The pattern of these correlations suggested that the effects might be modeled as age and gender relationships to the second-order factors. To investigate this, the hierarchical model specified in Model H5 was added to the analysis with age and gender. Relationships of age and gender to BSRI factors were specified as directed regression coefficients to factors and items. The specific item relationships to age and gender of model R2 were retained, but only the two second-order factors, Masculinity and Femininity, were allowed to have salient regressions upon the exogenous age and gender variables.

Table VIII. Factor Correlations of BSRI First-Order Factors with Age and Gender (Combined Sample)[a]

	Isolated factors model		Factors and items model	
Factor	Age	Gender	Age	Gender
IAF	.13[b]	.17[b]	.12[b]	.17[b]
DEC	.00	–.23[b]	.04	–.24[b]
SHY	.00	–.05	–.19[b]	–.04
S-S	.11	–.14[b]	.13[b]	–.14[b]
ANL	.01	–.16[b]	.01	–.16[b]
DOM	–.02	–.27[b]	–.07	–.20[b]
COM	.19[b]	.25[b]	.19[b]	.25[b]

[a]For abbreviations, see footnote [a] of Table III.
[b]$p < .01$.

This model, R4, did not fit as well as R2 (Table VII). There were some salient modification indices for age and gender-related paths. It did not appear, then, that the age and gender differences in the BSRI could be modeled as being fully mediated by global Masculinity and Femininity. In addition, the estimated path of age to Masculinity was not significantly different from zero.

We used the modification indices as a basis for adding additional regressions of first-order factors on age and gender. Specifically, we added (a) paths from age to Self-Sufficient, Shy, and Compassionate, and (b) a path from gender to Compassionate. These effects were admittedly small, but did improve the model fit. The nonsignificant regression of Masculinity on age was fixed to zero. The final Model R5 did not fit as well as R2, but was deemed an adequate solution given the high RNI and RFI indices.

Figure 2 shows the standardized regression coefficients from Model R5. Consistent with the factor correlations from the factor analysis model, the relationships of gender and age to the BSRI factors were small in magnitude. Masculinity yielded the largest standardized coefficient for gender (–.20). The partial correlation of Masculinity and Femininity, controlling for age and gender, was .40.

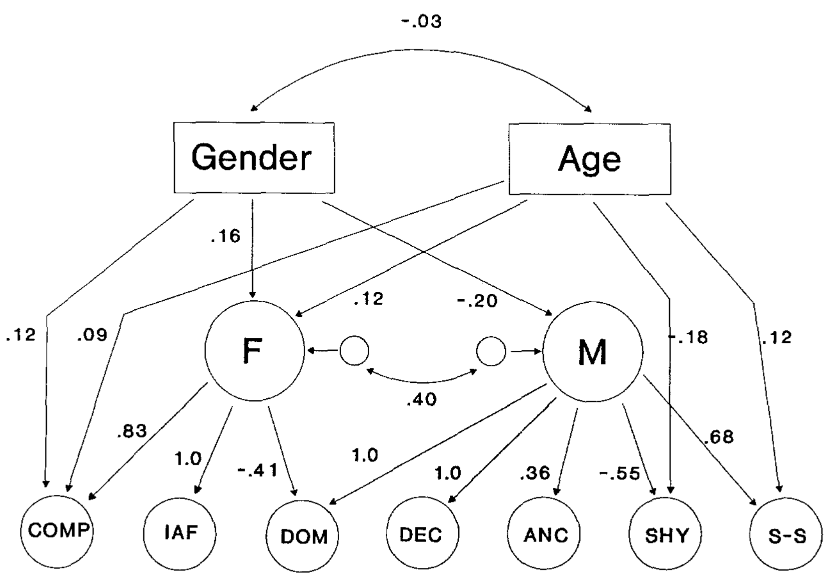

Fig. 2. Structural regression model of BSRI first- and second-order factors with age and gender as predictors. See footnote *a* of Table III for abbreviations.

DISCUSSION

Since the development of self-report gender role inventories, such as the BSRI, there have been mounting criticisms reflecting both methodological and theoretical concerns. The results of these analyses address several of these concerns regarding the validity, interpretation, and use of the BSRI in gender role research.

Factor Structure of the BSRI

First-Order (Item) Factors. The confirmatory factor analyses indicate a multidimensional factor structure, with varying degrees of relation among factors (Spence, 1984b) rather than a two-factor structure as originally proposed (Bem, 1974). Moreover, the present results appear capable of reconciling what has seemed to be rather discrepant findings in the literature regarding the factor structure of the BSRI.

The first-order factor analysis indicated a large number of BSRI item factors; the accepted CFA solution produced 7-item factors and 3 free-standing items. To be sure, several of these factors were highly intercorrelated. For example, we differentiated a Dominant factor, including items such as dominant and aggressive, from a Leadership/Assertiveness factor, although the two factors correlated highly with each other. Likewise, we identified two highly correlated factors relating to the BSRI femininity items.

The chief difference between the item factor structure identified here and past analyses of the BSRI is that 3 items normally assigned to the BSRI Femininity scale (Shy, Timid, Yielding) form a separate factor that correlates most highly but negatively with the Assertive and Dominant factors. Thus, these items are best conceptualized as measuring a negative aspect of source traits typically associated with global masculinity. In contrast, the item factors most centrally related to the femininity item pool, Interpersonal Affect and Compassion, correlate positively with the Assertive and Dominant item factors (see also Marsh & Myers, 1986).

Assigning the items measuring shyness to a global Femininity scale will suppress a positive correlation between it and a global Masculinity scale. For example, the correlation between BSRI Masculinity and Femininity scales in our sample is –.04 when computed by using the original Bem scoring system. However, when Masculinity and Femininity were scaled using 22 masculinity items (with reverse scoring on the 3 shyness items) and only 14 femininity items (from the IAF and Compassion factors), the two scales have a correlation of .26 in the present sample.

The positive correlations between masculine-related and feminine-related item factors argue against the popular conception of masculinity and femininity as negatively correlated attributes, and certainly falsify the hypothesis that these factors, as measured by the BSRI, are bipolar opposites of a single dimension. Nevertheless, it should be noted that the positive correlation between Masculinity and Femininity, as measured by the BSRI, may be idiosyncratic to that instrument. Marsh, Antill, and Cunningham (1989) argue that the use of socially desirable items on the BSRI is more likely to produce positive correlations between Masculinity and Femininity than would be the case for scales with items balanced in relation to social desirability. In the case of items related to shyness, the effect works in the opposite direction to that noted by Marsh et al. (1989), only because these items have traditionally been associated with the femininity dimension instead of being conceptualized as negatively valenced masculinity items. Marsh et al.'s overall point is still well taken: the correlations among scales measuring masculinity and femininity will be influenced by the valences of the set of gender-related source traits measured by the scale.

Second-Order Factor Structure. Although the item factor analysis rejected a two-factor solution for the BSRI, the concept of embedded Masculinity and Femininity factors is supported by the results from the hierarchical factor analysis. As suggested by Marsh and Myers (1986), global Masculinity and Femininity are probably best conceived as multifaceted dimensions of attributes that can be identified as second-order factors. We were able to successfully estimate Masculinity and Femininity in the hierarchical models, finding a positive correlation between the two factors.

The solution we found has some notable features that differ from that of Marsh and Myers (1986), however. First, Shy is related (negatively) to Masculinity not Femininity. Second, Dominant has a negative relationship to Femininity in addition to its relationship to Masculinity. Finally, the model did not fully account for all first-order factor covariances, and we detected the possibility of adding an additional second-order factor (related to the Self-Sufficient, Analytical, and Shy item factors). Although we opted to retain the Masculinity and Femininity solution, the possibility of an additional factor is interesting because it would support the suggestion by several previous studies that there are multiple masculinity-related factors present in the BSRI (e.g., Pedhazur & Tetenbaum, 1979).

One implication of fitting Masculinity as a second-order, rather than a first-order, factor is that it becomes possible to ask whether relationships of the BSRI to other variables (e.g., adjustment, coping strategies) are a function of higher order Masculinity and Femininity, or alternatively, are specific to first-order factors such as Self-Sufficient. Indeed, in a recent study cross-cultural differences on Masculinity were not detected using the standard BSRI global Masculinity scale, but emerged when two masculinity subscales were differentiated (Norris, Blanchard-Fields, & Arguelles, 1993). The Self-Sufficient scale yielded different patterns of age and gender differences between Spain and the United States, whereas the Dominant/Decisive scale produced no cross-cultural differences.

Gender-Related Equivalence of BSRI Factor Structure. In addition to confirming the overall factor structure of the BSRI, the present analysis indicates that the item factor structure of the BSRI is invariant for males and females. The same seven factors were identified in both gender groups and were shown to have equivalent factor pattern matrices. The literature on comparative factor analysis indicates that, when factors have different mean levels across subpopulations (as is the case for the BSRI factors), one can expect invariance only at the level of equivalent factor loadings (see Mulaik, 1972). Thus the present finding of equivalent numbers of factors and of gender-related invariance in factor loadings suggests the same basic item factors are present in both male and female subpopulations. It appears, then, that the gender differences in the BSRI factor structure reported in previous research (Pedhazur & Tetenbaum, 1979; Windle & Sinnott, 1985) may have been an artifact of the exploratory factor analytic techniques used in those studies. Gender differences in the BSRI factors appear to be best characterized as differences in mean levels on equivalently defined (but gender-correlated) attributes.

There were, however, some gender differences in the correlations among item factors. The major differences involved correlations of self-rated masculinity and femininity with other BSRI factors (see below). The differences in correlations among other factors were relatively small and, given the gender differences in means, do not contradict the conclusion that the factor structure of the BSRI is invariant between gender groups (Mulaik, 1972). The basis for this argument is Meredith's (1964) work on selection and factorial invariance. He showed that, even if a common factor model held for all members of a population, selection into subgroups differing in mean levels on the factor scores would result in nonequivalence in factor covariances across the subgroups. Clearly males and females differ on the gender-related attributes measured by the BSRI, and hence we can expect some impact of these selection processes on the factor covariances within each gender.

Gender, Masculinity, and Femininity

One somewhat surprising outcome of the present study was the limited size of the relationship between gender and the BSRI first-order and second-order factors. The structural regression analyses showed that gender did significantly predict the BSRI factors, but the standardized regression coefficients were generally weak in magnitude. The results are consistent with Pedhazur and Tetenbaum (1979), who found that the Masculine and Feminine items accounted for the bulk of the gender-related variance in the BSRI.

Given that there are many studies in the literature reporting gender differences on the BSRI, it is appropriate to ask whether the results reported here are atypical, perhaps reflecting the age-heterogeneity of the present sample. It appears, instead, that the typical finding of gender differences on the BSRI reflects the common practice of computing statistical significance tests on mean gender differences (usually analysis of variance *F* tests) without necessarily attending to the magnitude of the significant effects. For example, Feldman et al. (1981) reported significant large-sample gender differences in means on a number of BSRI subscales. We converted the mean differences reported by Feldman et al. (1981) to an eta^2 statistic, however, and found that the differences generally accounted for less than 4% of the variance in the BSRI subscales—effect sizes comparable with those reported here. Thus gender differences in attributes measured by the BSRI are robust, in the sense that they are replicable, but they appear to be relatively modest in magnitude.

However, mean gender differences in BSRI subscales may not tell the full story—or even the most interesting story—regarding gender differences in gender role orientation as measured by the BSRI. Biological gender moderates (in the sense of statistical interaction) the relationship of self-rated masculinity and femininity with the other BSRI item factors. Gender-consistent ratings correlate higher with self-ratings on positive gender-related attributes for each gender group. Thus, females' ratings on the Feminine item correlate positively with such factors as Interpersonal Affect and Compassionate and males' ratings on the Masculine item correlate positively with Decisiveness and Dominant. However, a markedly different pattern of correlations appears when gender-inconsistent ratings are examined, particularly when factors reflecting feminine content, i.e., Interpersonal Affect and Compassion, are correlated with the Masculine item. In this case, males' masculinity ratings correlate positively, and significantly, with Interpersonal Affect and Compassionate, whereas females' masculinity ratings correlate negatively with these factors. Thus, a male who rates himself as highly masculine will also tend to rate himself as interpersonally sensitive. However, for females the more traditionally expected negative relationship between masculinity and interpersonal sensitivity is evidenced.

This pattern of effects has not, to our knowledge, been previously reported in the literature on the BSRI. The reason is probably a consequence of the tendency for studies to (a) to combine males and females into a gender-aggregated sample, and (b) compute a bipolar M-F factor from the Masculine and Feminine items (Marsh & Myers, 1986). Our study shows that analysis of a gender-aggregated sample will accentuate the loadings on the bipolar M-F factor (see also Pedhazur & Tetenbaum, 1979), but it also questions the adequacy of the bipolar M-F factor in gender-segregated samples. Clearly the interactive relationships associated with self-rated masculinity and femininity cannot be detected if the two items are combined into a single variable.

Gender differences in the relationship between self-rated masculinity and femininity and other BSRI item factors raise an interesting set of issues regarding how an individual perceives her/his gender role orientation. That males tend to see themselves as both masculine and interpersonally sensitive, whereas females do not, suggests that there is a qualitative difference in the way females and males construe the meaning of masculinity and femininity. This finding appears to be consistent with the multiplicity model of gender identity espoused by Ashmore and his colleagues. Ashmore (1990) describes a prototypical example from his study of gender-related attributes as applied to social relationships. The individual male in this example saw himself as "masculine" with the important individuals in his life. However, when asked what "masculine" means to him, he responded: "I feel responsible to help other people," a response that appears consistent with the present finding that masculine males see themselves as interpersonally sensitive.

Assessment of masculinity and femininity using self-reports on the BSRI may conceptualize an individual's gender role orientation strictly from the perspective of the test developer. Instead, these findings suggest the need to study gender role orientation using a method that takes into consideration the individual's role in structuring the meaning of gender identity. In other words, gender identity must be defined in terms of the individual rather than an experimenter-imposed construction of a scale. Future research will need to address such questions as follows: how is interpersonal sensitivity defined by males and females? Is it an equivalent constructs for both genders? How is interpersonal sensitivity behaviorally manifested by females and males, and in what contexts?

The pattern of relationships between biological gender and the BSRI items casts further doubt on the validity of gender-typed classifications (masculine, feminine, androgynous) on the basis of BSRI self-ratings. The common practice of using median splits on the BSRI Masculinity and Femininity subscales has been extensively critiqued elsewhere (e.g., Cook, 1985; Pedhazur & Tetenbaum, 1979; Spence, 1984b). The limited magnitude of relationship between biological gender and BSRI factors, suggests the potential for misclassification due to measurement error may be relatively high. More importantly, the finding of (a) invariance in BSRI item factor structure, combined with (b) gender differences in relationships of self-rated masculinity and femininity to BSRI item factors, suggests that BSRI item factors derived from self-ratings may be insensitive to crucial aspects of gender identity and internalized gender role stereotypes.

Developmental Implications for Gender-Related Attributes

We also explored developmental differences in BSRI attributes across the adult life span. The structural regression analyses indicate that, indeed, there are age differences in Interpersonal Sensitivity over the adult life span, although—as in the case of gender differences—the effects are relatively small in magnitude. Moreover, we detected age differences on specific BSRI items above and beyond the relationship of age to the BSRI item factors. This latter finding is important, for it suggests that traditional BSRI scale scores for masculinity and femininity may overestimate age differences by combining item-specific differences with differences more appropriately identified with the item factors per se. The structural regression models suggest that, controlling for these item-specific effects, there is a weak but positive association of chronological age with Interpersonal Sensitivity—older persons report themselves to be more sensitive than younger persons. Given our cross-sectional sample, this difference could represent either developmental change or generational differences. In addition, this correlation (between age and interpersonal sensitivity) could account for findings reported by Hyde and her colleagues (Hyde & Phillis, 1979; Hyde, Krajnik, & Skuldt-Niederberger, 1991) that androgyny increases for men in older age groups. They used the traditional median split method of BSRI gender role classification. Although their interpretation argued for a qualitative shift in males' gender role orientation across the life span, our analysis suggests a rather different interpretation as being equally likely: continuity in gender role orientation, coupled with small developmental increases for both males and females in mean levels of sensitivity and interpersonal affiliation.

Conclusion

Based on LISREL analyses, the BSRI most appropriately assesses two multifaceted factors, global Masculinity and Femininity, with the possibility of an additional second-order masculinity factor. These attributes can be differentiated from the assessment of biological gender identity (i.e., ratings on the BSRI Masculine and Feminine items). In fact, the differential relationship for females and males between self-rated masculinity and femininity and BSRI factors suggests the need for additional methods of describing individuals' gender role conceptualizations. Like Spence (1984a), we argue that the BSRI and similar scales measure a set of gender-related personality attributes. From this point of view, masculinity and femininity (as well as androgyny) are best interpreted as a profile or pattern of scores on a set of attributes known to correlate with biological gender. Although this point may seem obvious to those familiar with the original conceptualization of gender typing and gender roles, it seems to have become obscured in the literature on the factor structure of gender role inventories like the BSRI. This interpretation is consistent with work demonstrating convergence between the BSRI and constructed indices of masculinity and femininity using scales from the Jackson Personality Research Form (Berzins et al., 1978).

Given that masculinity and femininity might be defined as conforming to a gender-typed profile of scores on gender-related attributes, one can actually argue that the number of factors present in the BSRI is not the critical issue in determining the validity of ascribing traditional gender typing on the basis of scores on the BSRI. One can wonder, nevertheless, about the content validity of the dimensions used to generate the profile. Personality attributes (e.g., trait anxiety) known to be related to gender are not included in the BSRI (Pedhazur & Tetenbaum, 1984; Spence, 1984b), and aspects of emotionality (i.e., emotional regulation) that may be core features of gender stereotypes (Labouvie-Vief, DeVoe, & Bulka, 1989; Morawski, 1987) may be inadequately represented by the BSRI's Interpersonal Sensitivity factor.

Moreover, there may be dimensions beyond personality variables that characterize an individual's gender role orientation. Ashmore and his colleagues (Ashmore, 1990; Ashmore & Del Boca, 1986) suggest that a simple gender-as-a-personality-variable approach is too simplistic. Instead, gender role orientation consists of loosely connected multiple components such as gender-related personality attributes, stereotypes, attitudes, behavior, social relationships, interests, and abilities, etc., that must be considered in a social context (Ashmore, 1990). Therefore, future research needs to address the question of important variables for representing an individual's gender role orientation that have been omitted from gender role inventories such as the BSRI.

REFERENCES

Ashmore, R. D. (1990). Sex, gender, and the individual. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research.* New York: Guilford Press.

Ashmore, R. D., & Del Boca, F. K. (Eds.). (1986). *Social psychology of male-female relations: A critical analysis of central concepts.* Orlando, FL: Academic Press.
Bem, S. L. (1974). Measuring psychological androgyny. *Journal of Consulting and Clinical Psychology, 41,* 155-162.
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. (1984). Androgyny and gender schema theory: A conceptual and empirical integration. In T. B. Sondregger (Ed.), *Nebraska Symposium on Motivation 1984: Psychology and gender* (Vol. 32). Lincoln: University of Nebraska Press.
Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.
Berzins, J. I., Welling, M. A., & Wetter, R. E. (1978). A new measure of psychological androgyny based on the Personality Research Form. *Journal of Consulting and Clinical Psychology, 46,* 126-138.
Bledsoe, J. C. (1983). Factorial validity of the Bem Sex-Role Inventory. *Perceptual and Motor Skills, 56,* 55-58.
Bollen, K. A., & Long, J. S. (Eds.) (1993). *Testing structural equation models.* Newbury Park, CA: Sage.
Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445-455.
Carlsson, M. (1981). Note on factor structure of the Bem Sex-Role Inventory. *Scandinavian Journal of Psychology, 22,* 123-127.
Collins, M., Waters, C. W., & Waters, L. K. (1979). Factor analysis of sex-typed items from the Bem Sex Role Inventory: A replication. *Psychological Reports, 44,* 517-518.
Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80,* 389-407.
Cook, E. P. (1985). *Psychological androgyny.* New York: Pergamon Press.
Feldman, S. S., Biringen, Z. C., & Nash, S. C. (1981). Fluctuations of sex-related self-attributions as a function of stage of family life cycle. *Developmental Psychology, 17,* 24 35.
Gaa, J. P., Liberman, D., & Edwards, T. A. (1979). A comparative factor analysis of the Bem Sex Role Inventory and the Personal Attributes Questionnaire. *Journal of Clinical Psychology, 35,* 592-598.
Gross, R., Batlis, N., Small, A., & Erdwins, C. (1979). Factor structure of the Bem Sex-Role Inventory and the Personal Attributes Questionnaire. *Journal of Consulting and Clinical Psychology, 47,* 1122-1124.
Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore, MD: Johns Hopkins University Press.
Hertzog, C. (1989). Using confirmatory factor analysis for scale development and validation. In M. P. Lawton & A. R. Herzog (Eds.), *Special research methods for gerontology.* New York: Baywood Press.
Hertzog, C. (1990). On the utility of structural equation models for developmental research. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 10). Hillsdale, NJ: Lawrence Erlbaum Associates.
Hertzog, C., Van Alstine, J., Usala, P. D., Hultsch, D. F., & Dixon, R. A. (1990). Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment, 2,* 64-72.
Hiller, D. V., & Philliber, W. W. (1985). Internal consistency and correlates of the Bem Sex Role Inventory. *Social Psychology Quarterly, 48,* 373-380.
Hyde, J. S., & Phillis, D. E. (1979). Androgyny across the lifespan. *Developmental Psychology, 15,* 334-336.
Hyde, J. S., Krajnik, M., & Skuldt-Niederberger, K. (1991). Androgyny across the life-span: A replication and longitudinal follow-up. *Developmental Psychology, 27,* 516-519.
Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409-426.

Jöreskog, K. G., & Sorbom, D. (1984). *LISREL VI user's guide.* Mooresville, IN: Scientific Software.

Labouvie-Vief, G., DeVoe, M.,& Bulka, D. (1987). Speaking about feelings: Conceptions of emotion across the life span. *Psychology and Aging, 4,* 425-437.

Larsen, R. J., & Seidman, E. (1986). Gender schema theory and sex role inventories: Some conceptual and psychometric considerations. *Journal of Personality and Social Psychology, 50,* 205-211.

Marsh, H. W. (1985). The structure of masculinity/femininity: An application of confirmatory factor analysis to higher order factor structures and factorial invariance. *Multivariate Behavioral Research, 20,* 427-449.

Marsh, H. W., & Hocevar D. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. *Psychological Bulletin, 97,* 562-582.

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

Marsh, H. W., Antill, J. K., & Cunningham, J. D. (1989). Masculinity and femininity: A bipolar construct and independent constructs. *Journal of Personality, 57,* 625-661.

Maznah, I., & Choo, P. F. (1986). The factor structure of the Bem Sex-Role Inventory (BSRI). *International Journal of Psychology, 21,* 31-41.

McDonald, R. P. (1978). A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 31,* 59-72.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107,* 247-255.

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika, 29,* 187-206.

Morawski, J. G. (1987). The troubled quest for masculinity, femininity, and androgyny. In P. Shaver & C. Hendrick (Eds.), *Review of personality and social psychology: Vol. 7. Sex and gender.* Beverly Hills, CA: Sage.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, C. D. (1989). An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin, 105,* 430-445.

Norris, L., Blanchard-Fields, F., & Arguelles, A. (1993). *A cross cultural study of gender roles: Attitudes toward women from an adult developmental perspective.* Unpublished manuscript.

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.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23,* 51-67.

Ruch, L. O. (1984). Dimensionality of the Bem Sex Role Inventory: A multidimensional analysis. *Sex Roles, 10,* 99-117.

Sassenrath, J. M., & Younge, G. D. (1979). The Bem Sex-Role Inventory reexamined. *Psychological Reports, 45,* 935-941.

Schmitt, B. H., & Millard, R. T. (1988). Construct validity of the Bem Sex Role Inventory: Does the Bem Sex Role Inventory distinguish between gender-schematic and gender aschematic individuals? *Sex Roles, 19,* 581-588.

Spence, J. T. (1984a). Gender Identity and its implications for the concepts of masculinity and femininity. In T. B. Sondregger (Ed.), *Nebraska Symposium on Motivation 1984: Psychology and gender* (Vol. 32). Lincoln: University of Nebraska Press.

Spence, J. T. (1984b). Masculinity, femininity, and gender-related traits: A conceptual analysis and critique of current research. In B. Maher & W. Maher (Eds.), *Progress in experimental personality research: Vol. 13. Normal personality processes.* Orlando, FL: Academic Press.

Tanaka, J. S., & Huba, G. J. (1984). Structure of psychological distress: Testing confirmatory hierarchical models. *Journal of Consulting and Clinical Psychology, 52,* 719-721.

Thompson, B., & Melancon, J. G. (1986). Factor structure of the Bem Sex Role Inventory. *Measurement and Evaluation in Counseling and Development, 19,* 77-83.

Waters, L. K., & Popovich, M. (1986). Factor analysis of sex-typed items from the Bem Sex Role Inventory: A multiple replication across time. *Psychological Reports, 59,* 1323-1326.

Waters, C. W., Waters, L. K., & Pincus, S. (1977). Factor analysis of masculine and feminine sex-typed items from the Bem Sex Role Inventory. *Psychological Reports, 40,* 56-57.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37,* 395-412.

Wiggins, J. S., & Pincus, A. L. (1992). Personality: Structure and assessment. *Annual Review of Psychology, 43,* 505-529.

Windle, M. (1986). Methodological and statistical considerations in the measurement of sex roles and aging. In J. D. Sinnott (Ed.), *Sex roles and aging: Theory and research from a systems perspective.* New York: Karger.

Windle, & Sinnott (1985). Psychometric study of the Bem Sex Role Inventory with an older adult sample. *Journal of Gerontology, 40,* 336-343.