The Spanish Journal of Psychology
2010, Vol. 13 No. 2, 1000-1009

ISSN 1138-7416

Do the BSRI and PAQ Really Measure Masculinity and Femininity?

Juan Fernández and Mª Teresa Coello

Universidad Complutense (Spain)

The two most used instruments to assess masculinity (M) and femininity (F) are the Bem Sex Role Inventory (BSRI) and the Personality Attributes Questionnaire (PAQ). Two hypotheses will be tested: a) multidimensionality versus bidimensionality, and b) to what extent the two instruments, elaborated to measure the same constructs, classify subjects in the same way. Participants were 420 high school students, 198 women and 222 men, aged 12-15 years. Exploratory factor analysis and internal consistency analysis were carried out and log-linear models were tested. The data support a) the multidimensionality of both instruments and b) the lack of full concordance in the classification of persons according to the fourfold typology. Implications of the results are discussed regarding the supposed theory behind instrumentality/expressiveness and masculinity/femininity, as well as for the use of both instruments to classify different subjects into the four distinct types.

Keywords: androgyny, masculinity, femininity, BSRI, PAQ, log-linear model.

Los dos instrumentos más utilizados para valorar masculinidad y feminidad son el Bem Sex Role Inventory (BSRI) y el Personality Attributes Questionnaire (PAQ). Se pondrá a prueba la hipótesis de la multidimensionalidad frente a la de la bidimensionalidad. A su vez, se tratará de verificar hasta qué punto ambos instrumentos, que dicen medir lo mismo, clasifican a los sujetos de igual forma. Los participantes fueron 420 estudiantes de secundaria, 198 mujeres y 222 varones, de entre 12 y 15 años. Se llevaron a cabo análisis factoriales exploratorios y de consistencia interna y se pusieron a prueba modelos lineal-logarítmicos. Los datos apoyan: a) la multidimensionalidad de ambos instrumentos y b) la falta de plena concordancia en la clasificación de las personas en función de la cuádruple tipología. Se analizan las implicaciones de los resultados tanto para la supuesta teoría subyacente –instrumentalidad/expresividad, masculinidad/feminidad– como para la utilización de ambos instrumentos a la hora de clasificar a los sujetos en cuatro tipos distintos.

Palabras clave: androginia, masculinidad, feminidad, BSRI, PAQ, modelo lineal-logarítmico.

Correspondence concerning this article should be addressed to Juan Fernández. Departamento de Psicología Evolutiva y de la Educación. Facultad de Psicología. Universidad Complutense de Madrid. Campus de Somosaguas. 28223 Madrid. (Spain). E-mail: jfernandez@psi.ucm.es. Web page: http://sites.google.com/site/jfsprofile/

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

It is widely known that in the last quarter of the 20th century, within the study of psychology, the instruments most often used to assess the concepts of masculinity (M) and femininity (F) have been the Bem Sex Role Inventory -BSRI- and the Personality Attributes Questionnaire -PAQ- (Beere, 1990; Lenney, 1991). Their common denominator is that they both involve two independent constructs, which lies in clear opposition to the bipolar continuum, which had been the predominant concept of the previous three quarters of the century (Gough, 1952; Hathaway & McKinley, 1943; Strong, 1936; Terman & Miles, 1936). Perhaps the article most relevant to this conceptual change, from bipolar to multidimensional, and from there to two independent dimensions, was that of Constantinople (1973). During the 1970's, the mainstream was bi-dimensionality, due to several theoretical convergences that had been taking hold since the 1950s (Bakan, 1966; Koestler, 1967, 1978; Parsons & Bales, 1955). These theories sparked the development of various instruments such as the BSRI and the PAQ (Baucom, 1976; Bem, 1974; Berzins, Welling, & Wetter, 1978; Heilbrum, 1976; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974, 1975). The force of this idea was so great that some authors even tried to use the old scales, theoretically based on the bipolar continuum, to assess independent bi-dimensionality (Woo & Oei, 2008). Furthermore, this perspective, which began in the United States, quickly crossed borders and expanded throughout the world, including Spanish-speaking countries (Agbayani & Min, 2007; Colley, Mulhern, Maltby, & Wood, 2009; Díaz-Loving, Rocha, & Rivera, 2004; Fernández, 1983; Kaschak & Sharratt, 1983; Leung & Moore, 2003; Peng, 2006). Afterwards, the possible relationship between the M and F dimensions (independent domains) began to be studied, as well as new dimensions (androgyny in particular), using a considerable number of psychological characteristics. This type of research continues today: M and F are related to autobiographical memory, moral reasoning, sexual behavior, social cognition, etc. (Ely & Ryan, 2008; Fink, Brewer, Fehl, & Neave, 2007; Kracher & Marble, 2008; Wood, Heitmiller, Andreasen, & Nopoulos, 2008).

During the last thirty years of research, since the appearance of these new instruments, they as well as their underlying assumptions have been highly criticized (Bem, 1979; Choi, Fuqua, & Newman, 2008; Pedhazur & Tetenbaum, 1979). This criticism assumes that conceiving M and F as two independent dimensions weakens both concepts, calling for a multidimensional view (Choi, Fuqua & Newman, 2006; Constantinople, 1973; Lippa, 2005; Marsh, 1985; Signorella, 1999; Spence, 1993). If we focus on the instruments, particularly the BSRI, the criticism –which is based on its factorial validity and the possible meanings of M and F– has been abundant (Brems & Johnson, 1990; Choi & Fuqua, 2003; Fernández, Quiroga, Del Olmo, & Rodríguez, 2007; Pedhazur & Tetenbaum, 1979; Marsh & Myers, 1986; Uleman & Weston, 1986; Wong, McCreary, & Duffy, 1990). Along with these criticisms, other studies have been carried out to support the theoretical basis of independent bidimensionality, as well as the validity of the instruments used to assess it. At times, it has merely been suggested that certain items be removed from the instrument whose present-day estimation would not be the same as in the 1970s (Auster & Ohm, 2000; Harris, 1994; Oswald, 2004).

In this study, we come from a theoretical approach that departs from some of the dominant concepts of our time (Fernández et al., 2007). First of all, we deem that there is no agreement between what people say they understand as M and F and the items selected from the BSRI and the PAQ to assess those constructs (Lippa, 2005; Myers & Gonda, 1982; Twenge, 1999). In fact, it would be interesting to check how frequently one of the items on the scale of masculinity (*masculine*) and one on the scale of femininity (*feminine*) of the BSRI constitute a bipolar factor that has little to nothing to do with the rest, neither in structure (bipolar), nor in correlation (low correlations with almost all of them) (Fernández et al., 2007).

Second, we assume that the mid-century theories that served as inspiration to these instruments were too general and ambiguous to be useful in this day and age, although in other contexts, they continue to be the object of in-depth analysis (Diekman & Eagly, 2000; Fiske, Cuddy, Glick, & Xu, 2002). We refer to concepts such as instrumentality (I) and expressiveness (E), those of agency and communion, and tendencies of self-assertiveness and integration (Bakan, 1966; Koestler, 1967, 1978; Parsons & Bales, 1955). The common denominator behind these concepts is rooted in considering family, or any small group of people in general, as an entity in which dual-leadership is wielded (fathers/mothers; men/women): One leader, the father/man, tries to ensure the family adequately and efficiently fulfills concrete societal objectives (external objectives of execution) and another, the mother/woman, worries about the cohesion of and positive relationships within this small group of people.

Third, we understand that selecting certain items (everything that is more desirable for one sex than for the other, in a given historical moment and society, specifically American society) allows one to predict with a high level of probability the appearance of many more factors than two, in the same way that happened with the M/F scales developed during the first half of the 20th century (Fernández, 1983).

Finally, we state that both instruments –which have been refined, particularly in their reduced versions– overlap considerably –in terms of the constructs of instrumentality and expressiveness (Good, Wallace, & Borst, 1994; Spence, 1991). Nevertheless, the two are not simply interchangeable when classifying individuals into the ever famous four-fold typology of androgynous, masculine, feminine and undifferentiated individuals.

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

In light of this, the hypotheses to be tested in this study of secondary school students will be essentially two: a) Both instruments –the BSRI and the PAQ– will turn out to be more multi-factorial than bi-factorial, contrary to the assumption upon which they were built; b) In spite of their overlap (both instruments are supposedly assessing I and E), which would logically predict a statistically significant association, the classification of participants into the four-fold typology will not be satisfactory. That is to say, we hypothesize that the frequency distribution of the four-fold typology (androgynous, feminine, masculine and undifferentiated) will be different for both instruments.

Method

Participants

In this study participated 420 students (in the first three years of Compulsory Secondary Education: 30% first-years, 32.3% second-years and 37.7% third-years). They all attend private schools in northern Madrid, and they range in age from 12 to 15 years old: 26.9% 12 years old, 29.3% 13, 34.8% 14 and 9% 15. Of them, 198 (47.1%) are women and 222 (52.9%) are men. Within each school, all students from each class were taken.

Instruments

Though the majority of research studies have used the 40-item version of the BSRI, 20 masculine and 20 feminine (Holt & Ellis, 1998; Konrad & Harris, 2002; Maznah & Choo, 1986), this study used an abbreviated version. These versions tend to show psychometric properties as well as or better than the original, so using such a version is advisable (Campbell, Gillaspy, & Thompson, 1997).

The abbreviated version used in this study was the one tested by Mateo and Fernández (1991) on university students. It is made up of only twelve elements: six assessing M (defends own beliefs, strong personality, has leadership abilities, makes decisions easily, dominant, acts as a leader) and six assessing F (affectionate, sympathetic, sensitive to needs of others, warm, tender, gentle). For the different participant groups (women and men together, men and women separately), the values of the coefficient of internal consistency (Carmines & Zeller, 1979) never fell below .83 and never exceeded .94. As for the variance accounted for by the factorial structures that were considered, this was around 58% in all cases, and there was a multi-factorial configuration. Subjects assessed themselves on each item according to a 7-point Likert scale, *1* signifying that the content of the item did not reflect him or her, and *7* meaning the item totally reflected him or her. The rest of the numbers on the scale represent intermediate values. The items were translated into Spanish by one of the authors in the early 1980s, and they were later translated into English by two bilingual people, one born and raised in the U.S. and the other in Spain. For the purposes of this study, to make the two instruments' response scales as similar as possible, the range of responses was from *1* to *5*.

The second instrument employed was the Personal Attributes Questionnaire (PAQ) (Spence & Helmreich, 1978; Spence et al., 1974, 1975). A short version of only 16 items, eight that assess I (independent, active, competitive, makes decisions easily, never gives up easily, self confident, feels very superior and stands up well under pressure) and eight that assess E (emotional, able to devote self to others, gentle, helpful, kind, aware of feelings of others, understanding of others and warm) was used. Participants were asked to indicate, similarly to the BSRI, to what extent the content of each item represented them, according to a 5-point Likert scale. The PAQ was translated in the same way as the BSRI.

Procedure

After communicating the study's objectives to the principals and/or teachers at the schools and obtaining their permission and collaboration, one of the authors (female) administered the reduced versions of the BSRI (12 items) and the PAQ (16 items). The questionnaires were filled out during class. Before beginning the questionnaires, the students were informed of the voluntary nature of their participation and the guaranteed anonymity of their responses. They were asked to respond as truthfully as possible. While administering the questionnaires, each item was read aloud to the participants while they followed along on their own. Next, they were asked if they had any questions regarding the statement they had read. Once clarifications (if necessary) had been made, they responded to the item. Half of the groups of participants took the BSRI first and then the PAQ, and the other half took the questionnaires in the reverse order.

Data Analysis

An exploratory factor analysis (EFA) was performed to examine the hypothesis of the dimensionality of the two instruments employed. This was done using the matrix of polychoric correlations between pairs of variables, obtained from Prelis, LISREL8 (Jöreskog & Sorbom, 1998). To determine the reliability of the instruments was performed a Cronbach's alpha coefficient, an analysis of internal consintency.

In order to examine the level of agreement between the two instruments (BSRI and PAQ) in categorizing the participants into the four-fold typology, log-linear models were used. These models allow us to examine the type of association between two categorical variables within the context of repeated measures and paired data (Agresti,

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

1990), such as in the present study. Depending on the established restrictions, the models could be of greater or lesser complexity. The most complex one is the saturated model, where the observed frequencies match the expected frequencies, so this doesn`t provide any information of interest. At the other end of the spectrum is the independence model (I), which allows us to contrast the hypothesis of no association between the two variables (the BSRI and PAQ typologies).

To analyze dependence, we used a series of alternative models appropriate for the characteristics of our data. The models were: 1) a *quasi-independence* (QI) model (Goodman, 1968), whose assumption is that there is no association between the variables (BSRI and PAQ typologies), excluding the cells along the diagonal; 2) models that allow us to analyze the symmetrical components of a table. Two types of models are included in this last category: a) a *quasi-symmetry* (QS) model that assumes that both triangular halves of the table are symmetrical –excluding the main diagonal– but that does not assume marginal homogeneity, and b) the *symmetry* model (S), that assumes symmetry in both triangular halves of the table (upper and lower), apart from the principal diagonal. This implies that the marginal distributions of the four categories are equal for the two entries in the table (BSRI and PAQ typologies). The goodness of fit statistic used was the likelihood ratio, G^2 (Agresti, 1990; Vermunt, 1998), whose distribution follows a chi-square model. Its degrees of freedom depend on the size of the table (4 x 4 in this case) and on the parameters to be estimated in each model. The program used to fit the log-linear models was *l*EM (Vermunt, 1998).

In order to establish the four-fold typology, the theoretical mean was used instead of the empirical median or the median provided by the authors (and their collaborators) for the two instruments. We believe that by doing so, the data are more coherent within its theoretical framework (androgynous = high M and F; masculine = high M and low F; feminine = high F and low M; undifferentiated = low M and F) and more easily comparable to the diverse body of research carried out in different countries –the same theoretical and empirical yardstick was always used in measurement. That being said, for the BSRI, the theoretical mean is 18 (six items times 3, which is the theoretical mean of a scale ranging from 1 to 5, such as the BSRI in this study) and for the PAQ it is 24 (eight items times 3).

Results

To test the first hypothesis, an EFA on each instrument (BSRI and PAQ) was carried out. This study has satisfied the established criteria that the number of cases must be greater than or equal to the result of multiplying the number of items by 5 (Lewis, 1995), and that ($N – n – 1$) must be greater than or equal to 50, where N is the number of participants and n is the number of variables (Lawley & Maxwell, 1971).

BSRI

The value of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .80, indicating that the fit of the data to the factorial model may be considered acceptable. Bartlett’s test of sphericity yielded statistically significant

Table 1
Factor Loadings Matrix and Communality Indexes of BSRI, Principal Axis and Oblimin Rotation

	Factors			
Items	1	2	3	h^2
9. Warm (F)	**.81**	-.06	.13	.55
10. Tender (F)	**.76**	-.15	-.04	.66
2. Affectionate (F)	**.71**	-.01	.08	.51
12. Gentle (F)	**.56**	-.05	**.31**	.37
6. Sensitive to needs of others (F)	**.53**	-.04	.09	.28
4. Sympathetic (F)	**.49**	-.01	**.34**	.32
11. Acts as a leader (M)	-.05	**.86**	**.33**	.75
5. Has leadership abilities (M)	-.01	**.86**	**.44**	.75
8. Dominant (M)	-.03	**.74**	**.32**	.55
7. Makes decisions easily (M)	.00	.28	.26	.10
1. Defends own beliefs (M)	.10	**.33**	**.53**	.30
3. Strong personality (M)	-.04	**.36**	**.47**	.26

Note. Loadings ≥ .30 (bold typed), are statistically significant (N = 420, p = .05 and statistical power = .80).
F = feminine items on the BSRI; M = masculine items on the BSRI.

results, χ^2 (66, N = 420) = 1579.37, $p < .01$, suggesting that the data is adequate for a factor analysis. To determine the number of factors, Kaiser-Guttman's rule of interpreting those associated with eigenvalues greater than 1 was applied. Utilizing this criterion, the results of the principal-axis EFA, with *oblimin* rotation and δ = 0, reveal a multi-factorial solution. This is illustrated by the results displayed in Table 1; three factors were extracted that explain 58.11% of the total variance.

The first of these factors explains 25.91% of the total variance and shall be called "Expressiveness", since it only includes items of femininity/expressivity (F/E). The second factor extracted explains 23.68% of the variance, and refers to "Instrumentality", which is made up of masculinity/instrumentality (M/I) items. The third, which accounts for 8.52% of the variance, is basically an M/I factor, but without the purity of the first one, since significant weights were found for two F/E items (4 and 12).

It is worth noting that one item was not found to be significant for any of the factors (7), and that items 1, 3, 4, 5, 8, 11 and 12 had statistically significant weights for two factors. It is also important to mention that the communalities were low for the majority of items, especially for items 1, 3, 4, 6, 7 and 12.

The principal-axis EFA with *varimax* rotation and the principal components EFA (using *varimax* or *oblimin*, where δ = 0) provided the same three-factor solution.

The correlation between factors 1 (F/E) and 2 (M/I) is not statistically significant, nor is the one between factors 1 (F/E) and 3 (M/I). The correlation between factors 2 and 3, on the other hand, is statistically significant (r_{23} = .41, $p <. 01$), but both refer to masculinity/instrumentality.

Even though the factorial structure does not support the existence of the two, originally predicted factors, Cronbach's alpha coefficient was calculated for both the original scales (α = .77 for the F/E scale; α = .73 for M/I) and the factorial solution obtained in this study: α = .77 for the first factor, α = .75 for the second and, α = .67 for the third.

PAQ

The KMO measure (.77) and Bartlett's test of sphericity, χ^2 (120, N = 420) = 1338.20, $p < .01$, indicate that the data is adequate for performing a factor analysis.

The same type of EFA was performed as was for the BSRI scale and the same criteria for retaining factors were applied. The results of the principal axis factor analysis (loadings and communality) are included in Table 2. Four factors, that together explain 51.18% of the total variance, were obtained: factor 1, F/E, explains 21.80%, factor 2, M/I, explains 13.85%, factor 3, M/I, 8.12% and factor 4, F/E, 7.41%. These results further demonstrate the multidimensionality of this instrument.

As for the PAQ, we encountered items with statistically significant weights for multiple factors (9 and 13) as well as

Table 2
Factor Loadings Matrix and Communality Indexes of PAQ Principal Axis and Oblimin Rotation

	Factors				
Items	1	2	3	4	h^2
14. Understanding of others (F)	**.70**	.25	-.10	**.50**	.66
8. Kind (F)	**.67**	-.02	-.01	.00	.47
5. Gentle (F)	**.63**	.01	-.04	.05	.40
15. Warm (F)	**.61**	.22	-.17	**.37**	.49
6. Helpful (F)	**.54**	.05	.04	.06	.30
4. Able to devote self to others (F)	**.54**	.20	-.02	.23	.32
9. Aware of feelings of others(F)	**.51**	.13	-.05	**.36**	.34
12. Self confident (M)	.10	**.63**	.15	-.11	.48
10. Makes decisions easily (M)	.11	**.62**	.17	.05	.39
16. Stands up well under pressure (M)	.01	**.47**	.14	.24	.24
11. Never gives up easily (M)	.08	**.39**	.28	.19	.20
1. Independent (M)	-.05	.25	.15	.17	.09
7. Competitive (M)	-.15	.17	**.81**	.05	.67
13. Feel very superior (M)	-.24	**.32**	**.41**	.09	.27
3. Active (M)	.13	.15	**.40**	-.00	.19
2. Emocional (F)	.19	.03	-.00	**.48**	.25

Note. Loadings ≥ .30 (bold typed), are statistically significant (N = 420, p = .05 and statistical power = .80).
F = feminine items on the PAQ; M = masculine items on the PAQ.

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

one item (1) whose weight did not turn out to be significant for any factor. Also, as was the case with the BSRI, the communalities were low for all items except 7 and 14.

The factorial solutions found using principal axis, *varimax* rotation and principal components (with *varimax* or *oblimin* and $\delta = 0$) also provided four factors.

Cronbach's alpha coefficient of reliability was calculated for the original scales, and it was found to be $\alpha = .74$ for F/E and $\alpha = .57$ for M/I. The alpha values found for factors 1 to 4 were: $\alpha = .76$, $\alpha = .54$, $\alpha = .45$ and $\alpha = .63$, respectively.

The correlations did not turn out to be statistically significant for the following pairs of factors: 1 (F/E) and 2 (M/I), 1 (F/E) and 3 (M/I), 1 (F/E) and 4 (F/E), 3 (M/I) and 4 (F/E). On the other hand, statistically significant correlations were found between factors 2 and 3 ($r_{23} = .29$, $p < .05$) and factors 2 and 4 ($r_{24} = .21$, $p < .05$).

The Four-fold Typology

As indicated above, participants were then classified according to the four-fold typology for both instruments according to the scales of M/I and F/E, considering: a) 18 to be the theoretical mean for the BSRI, while it was 24 for the PAQ, and b) that participants with scores greater than or equal to 18 or 24 would be considered androgynous; participants with scores greater than or equal to the theoretical mean for M and below the theoretical mean for F were classified as masculine; those with scores greater than the theoretical mean for F and below it for M were categorized as feminine, and those whose scores fell below the estimated means for both M and F were considered undifferentiated. Table 3 displays the cross-classification obtained from the two instruments: BSRI and PAQ.

In order to determine the level of agreement between the two classifications, a series of log-linear models was fitted to the data displayed in Table 3. The results corresponding to these models are shown in Table 4.

First, the absolute independence model (I) was fitted to the data to test if the two gender schema classifications (BSRI and PAQ) are unrelated. The value of the goodness of fit statistic for the I model, G^2 (9, N = 420) = 158.33, $p < .01$, leads us to reject the hypothesis of independence, as was expected. Next, the quasi-independence model (QI) was fitted to the data. This model tests the hypothesis of non-association between variables, excluding cells along the main diagonal, which represent agreement in the two instruments' classification. That is to say that the QI model implies the diagonal cells fit perfectly (for which association is assumed) and independence for the rest of the cells. In other words, the classifications found in other cells (outside the diagonal) should be at random. Thus the QI model assumes that participants classified into any category of the

Table 3
Classification of Gender Typology on the BSRI by Gender Typology on the PAQ: frequencies

	Four-fold typology (PAQ)				
Four-fold typology (BSRI)	Androgynous (1)	Masculine (2)	Feminine (3)	Undifferentiated (4)	Total
Androgynous (1)	**149**	11	19	0	179
Masculine (2)	15	**17**	1	1	34
Feminine (3)	78	8	**81**	12	179
Undifferentiated (4)	6	6	9	**7**	28
Total	248	42	110	20	420

Table 4
Log-linear Models: Goodness of Fit Statistics

Model	G^2	*df*	*p*
Independence (I)	158.33	9	.00
Quasi-independence (QI)	30.78	5	.00
Difference I - QI	127.55	4	.00
Quasi-symmetry (QS)	3.81	3	.28
Symmetry (S)	58.04	6	.00
Difference S - QS	54.22	3	.00

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

BSRI`s four-fold typology have equal probability of being classified into any of the other three categories, outside the main diagonal, in the PAQ`s four-fold typology; the same happens with the PAQ categories in relation to the BSRI ones. The results obtained, $G^2(5) = 30.78$, $p < .01$, indicate that although an important decrease was produced in the value of G^2, this model showed poor fit to the data. The difference between the values of the G^2 statistic for the I and QI models, $G^2(4) = 127.55$, $p < .01$, leads us to reject the independence hypothesis outside the main diagonal and to conclude that there is an association in the corresponding cells. That is to say, there is an association between the two typologies outside of the diagonal: participants classified into one of the four categories of the BSRI`s four-fold typology do not have the same probability of being classified into any of the other three categories of the PAQ`s four-fold typology. The same applies for the PAQ`s four-fold typology in relation to the BSRI one.

Bearing in mind that the fit to the QI model was not satisfactory, we next explored the existence of an association beyond the diagonal. We examined the symmetrical components to determine whether or not there were similar patterns of association in the two triangular areas of the matrix outside the diagonal. It was for this purpose that Caussinus's (1966) quasi-symmetrical (QS) model was tested. This model adds, to the previous one, a group of parameters for the symmetrical cells, representing a symmetrical effect. These parameters have the same values for the cells whose rows and columns have permutated ($\lambda_{ij} = \lambda_{ji}$). The goodness of fit statistic for this model, $G^2(3) = 3.81$, $p > .05$), indicates an improvement with regard to the previous model and an acceptable fit to our data. Furthermore, this model is less restrictive than the symmetrical one and assumes symmetry in the frequencies of the two halves of the table -excluding the diagonal- but it does not presume there to be marginal homogeneity.

The most restrictive model is that of symmetry (S), which assumes marginal homogeneity as well as symmetry in the two triangular halves of the table. The QS model is a particular case of S model: non-marginal homogeneity is assumed. The fit for model S, $G^2(6) = 58.04$, $p < .01$, is worse than that of the QS model. The difference between the goodness of fit statistics for QS and S, G^2 $(3) = 54.22$, $p < .01$, allows us to reject the hypothesis of marginal homogeneity, which means that the two instruments (BSRI and PAQ) give different frequency distributions for the four-fold typology.

Thus, of the models employed, the one with the most acceptable fit to the data was the quasi-symmetrical one. From the analyses, it was inferred that: 1) the two classifications are associated, as can be expected of two instruments that attempt to measure the same constructs, and the majority of cases fall along the main diagonal; 2) the fourfold typology distribution for BSRI differs from the one obtained for PAQ since the marginal distributions of the cross-classification table (Table 3) differed significantly, and 3) the two triangular halves of the table show an association between the two four-fold typologies which indicates that "classification errors" on one instrument with respect to the other should not be considered as random. In this way, those classified as androgynous on the BSRI, for example, have a greater probability of being classified in the feminine category on the PAQ, and those classified as androgynous on the PAQ have a greater probability of being classified as feminine than masculine or undifferentiated on the BSRI.

Symmetrically, of those classified as feminine on the BSRI, less than 50% are classified in the same category on the PAQ. This leaves 79% of those falling outside the diagonal assigned to the androgynous category on the latter instrument. As for the participants classified as masculine by the BSRI, we have found that 50% are classified in other categories on the PAQ, with the highest frequency being classified as androgynous. More than 50% of those classified as masculine on the PAQ are found outside the diagonal, with the majority being assigned to the androgynous category of the BSRI. Similarly, of those classified as undifferentiated on the BSRI, the majority of those found outside the diagonal are categorized as feminine. The same occurs when using the PAQ as the point of reference.

Discussion

The results clearly support the hypothesis of multidimensionality as opposed to bi-dimensionality in both evaluation instruments: the BSRI and the PAQ. The first thing to note about these results is that they were collected from secondary school students, that is, with participants ranging in age from 12 to 15 years old. This aspect is rarely found in international literature, since the majority of research has been carried out with university students as participants, including those conducted in Spanish-speaking countries (Choi & Fuqua, 2003; Fernández et al., 2007). Another aspect to bear in mind is that multidimensionality appears even when abbreviated versions of the two instruments are used, which are presumed to be better at demonstrating one of the instrument´s most basic assumptions: orthogonal bi-dimensionality (Colley et al., 2009; Peng, 2006).

Multidimensionality has been corroborated in secondary school students, as it had been both nationally and internationally before in college students, with the complete BSRI and PAQ, as well as with their abbreviated versions (Choi et al., 2006, 2008; Fernández et al., 2007). It seems appropriate to reflect upon the implications of these convergent results (although they were infrequently drawn from representative samples, as was the case with this study) to determine the underlying model of the two instruments.

First of all, we are obliged to recognize the considerable difference between presuming two orthogonal factors and finding three or four 3 or 4, with unpredicted correlations.

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

How might this be interpreted? In light of this data, it seems that for the BSRI, there is a connection between the lack of statistically significant correlation between the first F/E factor and the second and third M/I factors, while at the same time, the statistically significant correlation between the second and third factors is predictable. In both cases, we are talking about M/I. Nevertheless, if we square that correlation, we find that although it has statistical significance, the two factors share very little in common (it does not even reach 17%). How might this data be interpreted according to the original model that did not assume more than one type of M/I?

The PAQ situation complicates matters even further. Even when the correlations between the first factor (F/E) and the second and third factors (M/I) were not of statistically significant (in line with the model's predictions), why would the correlation between the second and fourth factors not be statistically significant if they are both F/E? Similarly, the statistically significant correlation between the second and third factors made sense, since they were both M/I, although it would be convenient to reanalyze this finding when the square correlation coefficient is considered (around 8%). More difficult to explain is the similarity between the statistically significant correlation between the second and fourth factors –I/E– (.21) and the second and third factors, both M/I (.29).

One must add to these difficulties the low communalities on almost all items in the two instruments, and especially in the PAQ. Also note that the proportion of variance accounted for was not exactly high in either of the instruments: it did not reach 60%. Moreover, the values for internal consistency leave room for improvement when it comes to the items on the original reduced scale, and when the items that comprise each factor are analyzed. To all of this, it is worth adding that on both instruments, there was one item that did not show significant weight for any of the factors.

In light of these results, one might ask if it is worth continuing to perform confirmatory factor analyses for this type of instrument, as is often the case, particularly for the BSRI (Agbayani & Min, 2007; Choi et al., 2006, 2008; Colley et al., 2009). If multidimensionality is confirmed as a fact beyond contextual differences; if scientific explanations fail to make sense of the results of multidimensionality; if we keep in mind that the communality values are rather low for the majority of items; if the proportion of variance accounted for is not exactly high; then what might the results of confirmatory factor analyses provide, other than adding further confusion to this area, which still finds itself in need of a minimal theoretical foundation about what masculinity/instrumentality and femininity/expressiveness actually mean?

As for the second hypothesis, we have confirmed that a considerable overlap is produced in the classification of participants into each of the four categories (androgynous, feminine, undifferentiated and masculine), which is to be expected as both instruments were designed to evaluate dimensions that are supposedly the same (M/I and F/E). However, the level of discrepancy was still considerable, as becomes clear when we subtract from the margins each of the four values of the diagonal. As occurred in the interpretation of correlations (beyond mere statistical meaning), here too it is suggested that the BSRI and the PAQ differ too greatly to be able to be considered interchangeable in classifying people into the four-fold typology. Bearing this in mind, it is not difficult to infer the possibility of accumulated error: a) when research is carried out on the characteristics of one of these four groups, b) when these four categories relate to different psychological variables (moral development or sexual behavior, to cite only a few of those studied), and c) when these categories are used as dependent or independent variables. In all these cases (a, b, c), there are combinations and mixtures of subjects in each category ("false positives and negatives") according to the instrument of evaluation (Fink et al., 2007; Kracher & Marble, 2008).

Upon considering these results on the whole, beyond the two hypotheses proposed, what implications might be inferred from them? It must be stated assertively that masculinity/instrumentality and femininity/expressiveness still do not seem to be even minimally well-defined. In fact, the scales created during the first half of the 20th century were not founded on theory but rather, on empirical reasoning: All items selected differentially by men and women later went on to become part of the M and F scales. Here we find one of the possible reasons for its problems (Constantinople, 1973; Fernández, 1983).

During the second half of the century, certain authors believed they had found the theory they needed behind the concepts of instrumentality and expressiveness (Parsons & Bales, 1955), which they tried to materialize in the form of the so-called new scales of M and F (Baucom, 1976; Bem, 1974; Berzins et al., 1978; Heilbrum, 1976; Spence et al., 1974, 1975). As has been highlighted (this study is one example), the terms are too ambiguous to serve as the basis of good assessment instruments.

It seems that the time has come, at the beginning of the 21st century, to try and open up other pathways to allow new theories to emerge (very probably framed within the notion of multidimensionality) that would be capable of more clearly laying the groundwork for new instruments (Lipa, 2005). Along these lines, it is important to highlight the convenience of distinguishing between two domains, the complex reality of sex and the no less complex reality of gender (Fernández et al., 2007). Sexology, on the one hand, and genderology on the other, should perhaps be charged with offering some new, more consistent theoretical framework from which coherent groups of items might be derived.

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

References

Agbayani, P., & Min, J. W. (2007). Examining the validity of the Bem Sex Role Inventory for use with Filipino Americans using confirmatory factor analysis. *Journal of Ethnic & Cultural Diversity in Social Work, 15*, 55-80.

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Auster, C. J., & Ohm, S. C. (2000). Masculinity and femininity in contemporary American society: A reevaluation using the Bem Sex-Role Inventory. *Sex Roles, 43*, 499-528.

Bakan, D. (1966). *The duality of human existence*. Chicago, CA: Rand McNally.

Baucom, D. H. (1976). Independent masculinity and femininity scales on the California Psychological Inventory. *Journal of Consulting and Clinical Psychology, 44*, 876.

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

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

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

Berzins, J. I., Welling, M. A., & Wetter, R. E. (1978). A new measurement of psychological androgyny based on the Personality Research Form. *Journal of Consulting and Clinical Psychology, 46*, 126-138.

Brems, C., & Johnson, M. E. (1990). Reexamination of the Bem Sex-Role Inventory: The interpersonal BSRI. *Journal of Personality Assessment, 55*, 484-498.

Campbell, T., Gillaspy, J. A., Jr., & Thompson, B. (1997). The factor structure of the Bem Sex-Role Inventory (BSRI): Confirmatory analysis of long and short forms. *Educational and Psychological Measurement, 57*, 118-124.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. London: Sage.

Caussinus, H. (1966). Contribution a l'ànalyse statistique des tableaux de corrélation. *Annales de la Faculté de Sciences de l´Université de Toulouse*, *29* (anné 1965), 77-182.

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

Choi, N., Fuqua, D. R., & Newman, J. L. (2006). Hierarchical confirmatory factor analysis of the Bem Sex Role Inventory. *Educational and Psychological Measurement, 67*, 818-832.

Choi, N., Fuqua, D. R., & Newman, J. L. (2008). The Bem Sex-Role Inventory: Continuing theoretical problems. *Educational and Psychological Measurement, 68*, 881-900.

Colley, A., Mulhern, G., Maltby, J., & Wood, A. M. (2009). The short form BSRI: Instrumentality, expressiveness and gender associations among a United Kingdom sample. *Personality and Individual Differences, 46*, 384-387.

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

Díaz-Loving, R., Rocha, T. E., & Rivera, S. (2004). Elaboración, validación y estandarización de un inventario para evaluar las dimensiones atributivas de instrumentalidad y expresividad. *Revista Interamericana de Psicología, 38*, 263-276.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin, 26*, 1171-1188.

Ely, R., & Ryan, E. (2008). Remembering talk: Individual and gender differences in reported speech. *Memory, 16*, 395-409.

Fernández, J. (1983). *Nuevas perspectivas en la medida de la masculinidad y feminidad*. Madrid: Editorial de la Universidad Complutense.

Fernández, J., Quiroga, M. A., Del Olmo, I., & Rodríguez, A. (2007). Escalas de masculinidad y feminidad: estado actual de la cuestión. *Psicothema, 19*, 357-365.

Fink, B., Brewer, G., Fehl, K., & Neave, N. (2007). Instrumentality and lifetime number of sexual partners. *Personality and Individual Differences, 43*, 747-756.

Fiske, S. T., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Good, G. E., Wallace, D. L., & Borst, T. S. (1994). Masculinity research: A review and critique. *Applied and Preventive Psychology, 3*, 3-14.

Goodman, L. A. (1968). The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *Journal of the American Statistical Association*, *63*, 1091-1131.

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

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

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

Heilbrum, A. B. (1976). Measurement of masculine and feminine sex roles identities as independent dimensions. *Journal of Consulting and Clinical Psychology, 44*, 183-190.

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

Jöreskog, K. G., & Sorbom, D. (1998). *LISREL8*. Chicago, IL: Scientific Software International.

Kaschak, E., & Sharratt, S. (1983). A Latin American Sex Role Inventory. *Cross-Cultural Psychology Bulletin, 18*, 3-6.

Koestler, A. (1967). *The ghost in the machine*. London: Hutchinson.

Koestler, A. (1978). *Janus: A summing up*. New York: Vintage Books.

Konrad, A. M., & Harris, C. (2002). Desirability of the Bem Sex-Role Inventory for women and men: A comparison between African Americans and European Americans. *Sex Roles, 47*, 259-271.

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press

Kracher, B., & Marble, R. P. (2008). The significance of gender in predicting the cognitive moral development of business practitioners using the Sociomoral Reflection Objective Measure. *Journal of Business Ethics, 78*, 503-526.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd. ed.). London: Buterworths.

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* (pp. 573-660). San Diego, CA: Academic Press.

Leung, C., & Moore, S. (2003). Individual and cultural gender roles: A comparison of Anglo-Australians and Chinese in Australia. *Current Research in Social Psychology, 8*, 302-316.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human–Computer Interaction*, *7*, 57-78.

Lippa, R. A. (2005). *Gender, nature, and nurture* (2nd. Ed.). Mahwah, NJ: LEA.

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

Mateo, M. A., & Fernández, J. (1991). La dimensionalidad de los conceptos de masculinidad y feminidad. *Investigaciones Psicológicas, 9*, 95-116.

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

Myers, A., & Gonda, G. (1982). Utility of the masculinity-femininity construct: Comparison of traditional and androgyny approaches. *Journal of Personality and Social Psychology, 43*, 514-522.

Oswald, P. A. (2004). An examination of the current usefulness of the Bem Sex-Role Inventory. *Psychological Reports, 94*, 1331-1336.

Parsons, T., & Bales, R. F. (Eds.). (1955). *Family, socialization, and interaction process*. New York: Free Press.

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

Peng, T. K. (2006). Construct validation of the Bem Sex Role Inventory in Taiwan. *Sex Roles, 55*, 843-851.

Signorella, M. L. (1999). Multidimensionality of gender schemas: Implications for the development of gender-related characteristics. In W.B. Swann, Jr., J.H. Langlois, & L.A. Gilbert (Eds.), *Sexism and stereotypes in modern society. The gender science of Janet Taylor Spence* (pp. 107-126). Washington, D.C.: American Psychological Association.

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

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

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

Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex roles 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.

Strong, E. K. (1936). Interest of men and women. *Journal of Social Psychology, 7*, 49-67.

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

Twenge, J. M. (1999). Mapping gender. The multifactorial approach and the organization of gender-related attributes. *Psychology of Women Quarterly, 23*, 485-502.

Uleman, J. S., & Weston, M. (1986). Does the BSRI inventory sex roles? *Sex roles, 15*, 43-62.

Vermunt, J. K. (1998). *A general program for the analysis of categorial data*. Tilburg: Tilburg University.

Woo, M., & Oei, T. P. S. (2008). Empirical investigations of the MMPI Gender-Masculine and Gender-Feminine Scales. *Journal of Individual Differences, 29*, 1-10.

Wood, J. L., Heitmiller, D., Andreasen, N. C., & Nopoulos, P. (2008). Morphology of the ventral frontal cortex: Relationship to femininity and social cognition. *Cerebral Cortex, 18*, 534-540.

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

Received November 27, 2008
Revision received August 5, 2009
Accepted September 21, 2009

https://doi.org/10.1017/S113874160000264X Published online by Cambridge University Press