Bem Sex Role Inventory Validation in the International Mobility in Aging Study*

Tamer Ahmed,[1] Afshin Vafaei,[2] Emmanuelle Belanger,[3] Susan P. Phillips,[4] and Maria-Victoria Zunzunegui[5]

RÉSUMÉ

Cette étude, en utilisant différentes méthodes d'analyse des facteurs, a examiné la structure de mesure de l'Inventaire des rôles sexués de Bem (IRSB). La plupart des études antérieures sur la validité ont appliqué analyse factorielle exploratoire (AFE) d'examiner l'IRSB. Il s'agissait d'évaluer les propriétés psychométriques et la validité de la construction de la forme courte IRSB comprenant 12 articles dans un échantillon administré à 1,995 personnes âgées de la vague 1 de l'Initiative internationale de la mobilité en viellissement (IIMV). Nous avons utilisé l'alpha de Cronbach pour évaluer la fiabilité et la cohérence interne et une analyse factorielle confirmatoire (AFC) afin d'évaluer les propriétés psychometriques. AFE a révélé un modèle comprenant trois facteurs, qu'on a confirmé par l'AFC, puis ceci est comparé avec le modèle structurel initial de deux facteurs. Les résultats ont révélé qu'une solution à deux facteurs (instrumentalité-expression) a montré satisfaisante validité conceptuelle et un ajustement supérieur aux données, par rapport à la solution à trois facteurs. La solution à deux facteurs confirme différences attendues entre les sexes chez les personnes âgées. L'IRSB composé de 12 articles fournit un instrument bref, psychométrique et fiable dans les échantillons internationaux des personnes âgées.

ABSTRACT

This study investigated the measurement structure of the Bem Sex Role Inventory (BSRI) with different factor analysis methods. Most previous studies on validity applied exploratory factor analysis (EFA) to examine the BSRI. We aimed to assess the psychometric properties and construct validity of the 12-item short-form BSRI in a sample administered to 1,995 older adults from wave 1 of the International Mobility in Aging Study (IMIAS). We used Cronbach's alpha to assess internal consistency reliability and confirmatory factor analysis (CFA) to assess psychometric properties. EFA revealed a three-factor model, further confirmed by CFA and compared with the original two-factor structure model. Results revealed that a two-factor solution (instrumentality-expressiveness) has satisfactory construct validity and superior fit to data compared to the three-factor solution. The two-factor solution confirms expected gender differences in older adults. The 12-item BSRI provides a brief, psychometrically sound, and reliable instrument in international samples of older adults.

[1] Pharmacist, PhD Candidate, Université de Montréal

[2] Department of Public Health Sciences, Queen's University, Ontario

[3] Postdoctoral Fellow, Institut de recherche en santé publique, Université de Montréal

[4] School of Medicine, Queen's University

[5] Institut de recherché en santé publique, Université de Montréal

* We are grateful to all the older adults who have given their time and confidence to this research. We would also like to thank our universities and institutes for providing continuous support and our funding sources: The Canadian Institutes of Health Research (grant number AAM 108751) and the Institut de Santé Publique de l'Université de Montréal.

Tamer Ahmed conceptualized and performed the data analysis and wrote the manuscript. Vafaei Afshin, Belanger Emmanuelle, Phillips Susan, and Zunzunegui M.V. contributed to interpretation of results and revised the manuscript. In addition to the co-authors of this article in the International Mobility in Aging Study (IMIAS), the following researchers have contributed at different steps of the study conception, design, data collection, and analyses, listed by their academic affiliation: E.E. Freeman and G. Karna at the Université de Montréal; N. Deshpande, A. Garcia, B. Alvarado, and S. Philips at Queens University; C.L. Curcio at Universidad de Caldas and at Guerra Rat Universidade Federale do Rio Grande do Norte, and A. Ylli at the Albanian National Institute of Public Health.

Manuscript received: / manuscrit reçu : 18/05/15

Manuscript accepted: / manuscrit accepté : 17/01/16

Mots clés : vieillissement, IRSB, rôles de gendre, stéréotypes de gendre, validation de la structure

Keywords: aging, BSRI, gender roles, gender stereotypes, structure validation

La correspondance et les demandes de tire-à-part doivent être adressées à : / Correspondence and requests for offprints should be sent to:
Tamer Ahmed, Pharmacist, PhD Candidate of Epidemiology
Institut de recherche en santé publique
Université de Montréal
7101 Ave du Parc, salle 3111
Montréal, QC H3N 1X9
(tis.ahmed@umontreal.ca)

Although the original "home" for the concepts of gender identity and personality attributes was psychology, more recent research has indicated their relevance to broader health research. *Gender identity* is the extent to which individuals perceive themselves as masculine and/or feminine given what each of these social constructs means in a given context and regardless of the individual's biological sex (Burke, Stets, & Piroggood, 1988; Spence, 1985; Stets & Burke, 2000). Gender roles are socially and culturally assigned personality attributes and behaviours expected of women and men (Lindsey, 2005). Both gender identity and roles are not fixed and may change with changes in time and place (Perry & Bussey, 1979; Spence, 1984). Gender role orientation (GRO) or gender stereotypes are determined and constrained by social systems through widely accepted judgment or bias regarding expected masculine and feminine behaviours; however, not all men and women conform to these stereotypes (Annandale & Hunt, 1990). Individually held stereotypes are reflected in gender role beliefs regarding the appropriate role of men and women in society (Bazik, 2011; Eagly & Mladinic, 1989).

Within any society, some personality characteristics, behaviours, interests, and roles are thought of as masculine whereas others are feminine (Williams & Best, 1982). Traditionally, masculinity has been characterized by a cognitive focus on "getting the job done" and is related to instrumental/agentic behaviours and attributes such as aggressiveness, assertiveness, competitiveness, and independence. The traditional traits that typify femininity are more affective and include concern for the welfare of others and communal-expressive behaviours and attributes such as submissiveness, dependence, deference, cooperation, caring, and nurturing (Bem, 1974; Williams & Best, 1982). A cross-cultural comparison of 14 countries showed that in egalitarian, more economically and socially developed countries, behavioural differences between sexes and self-perceptions of men and women are less stereotypical in nature than in less affluent countries, but they do still exist (Williams & Best, 1990).

Historically, masculinity and femininity were conceptualized as opposite ends of a continuum and linked to biological sex (Gough, 1952). The congruence between sex and sex roles was challenged by Constantinople (1973) who proposed a two-dimensional concept of gender in which masculinity and femininity are not two ends of a single scale, but are instead independent characteristics that could co-exist in an individual. Sandra Bem (1974) built on this, developing the Bem Sex Role Inventory (BSRI) to measure masculinity and femininity. This tool relies upon an individual's endorsement of a series of adjectives which have been judged as culturally characteristic of either males or females in the United States, on a scale from 1 ("never or almost never true") to 7 ("always or almost always true"). In its current form, the BSRI includes four categories based on answers to a 60-item Likert-type scale which comprises three 20-item scales about Masculinity (BSRI-M), Femininity (BSRI-F), and Social Desirability. Individuals who score high on masculinity and low on femininity are classified as "masculine". Similarly, individuals are classified as "feminine" if they score high on femininity and low on masculinity. High scorers on both the masculinity and femininity scales are classified as "androgynous", whereas those scoring low on both are classified as "undifferentiated".

Bem has suggested and validated an abbreviated 30-item version of the original BSRI that consisted of three subscales with 10 items each. Items were omitted from the longer BSRI if they were redundant or showed low correlation with the corresponding scale. The masculine and feminine scales of the 30-item BSRI appear to be more internally consistent than those of the original BSRI (Bem, 1981). Later on, the 30-item BSRI and many other abbreviated versions were developed to arrive at norms suited with different cultural settings and populations (Katsurada & Sugihara, 1999; Mateo & Fernández, 1991; Özkan & Lajunen, 2005). These abbreviated versions of the BSRI have been widely used among students, but their construct validity and research findings using abbreviated scales among older adults are limited. Moreover, issues have been raised concerning its cross-cultural relevance and validity (Ballardreisch & Elton, 1992; Hoffman & Borders, 2001; Holt, 1998; Zhang, Norvilitis, & Jin, 2001).

The measurement structure of the BSRI has been investigated using various methods of factor analysis. However, most of these studies were limited to *exploratory factor*

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

analysis (EFA) among samples of university students. A meta-analysis of previous EFA studies indicated that, with regard to feminine items, a single factor was most frequently found, whereas with the masculine items there were two or three factors (Choi & Fuqua, 2003). This suggests that the concept of masculinity may be more complex than that proposed by Bem. Interestingly, findings from studies among populations of older adults in Spain and Brazil supported the original BSRI two-factor structure of masculinity–femininity (Carver, Vafaei, Guerra, Freire, & Phillips, 2013; Vafaei et al., 2014), but in both studies no statistically significant association was observed between BSRI and biological sex.

Using EFA to determine the validity of latent constructs or concepts (factor validity) raises methodological issues (Bollen, 1989). Implementation of *confirmatory factor analysis* (CFA) in validity studies is important because such methods test models that are falsifiable (Brown, 2006). The few existing CFA studies on the BSRI have yielded some contradictory results. For instance, among high school and university students from Australia, a two-factor model provided reasonable fit to the BSRI data (Marsh, 1985), whereas findings from two samples of young adults and another of middle-aged participants in Midwestern U.S. states suggested a three-factor model with one feminine factor and two complex masculine factors produced the best fit (Choi, Fuqua, & Newman, 2009). Interestingly, results obtained with a sample of students from the United Kingdom also demonstrated a three-factor structure, with a bipolar factor reflecting interpersonal sensitivity/dominance, a masculine factor reflecting personal agency, and a third factor reflecting interpersonal expressiveness (Colley, Mulhern, Maltby, & Wood, 2009). Among populations over age 60, it is unclear whether a two-factor model adequately fits the BSRI data (Windle & Sinnot, 1985). Secondary findings from both waves of the longitudinal Social Relations and Mental Health over the Life Course Study suggest that the two-factor solution proposed by Bem fits the data well, on the basis of a 40-item version used in the first wave and an abbreviated 22-item version in the second wave of the study (Sellars, 2008). In summary, an overview of previous research highlights ongoing ambiguities in measuring gender roles across cultures, among older populations and using shorter versions of the inventory.

In this study, we used a 12-item BSRI short form. This abbreviated version of BSRI was initially tested for validity in a sample of university students from Spain (Mateo & Fernández, 1991) and has been re-examined in two separate pilot studies of Spanish and Brazilian older-adult populations by means of EFA and shown as a valid tool to measure gender role stereotypes (Carver et al., 2013; Vafaei et al., 2014).

Our aim was to examine the psychometric properties, as well as convergent and discriminant validity of the 12-item BSRI short form (6 items for measuring femininity and 6 for masculinity), in an international sample of older adults via both exploratory and confirmatory factor analysis.

Methods

Participants

The study subjects were part of wave 1 of the International Mobility in Aging Study (IMIAS) project. IMIAS is a prospective cohort study of 1,995 non-institutionalized men and women aged between ages 65 and 74 (as of 2012) from five sites: Natal, Brazil (n = 402), Manizales, Colombia (n = 400), Tirana, Albania (n = 394), Kingston (n = 398), and Saint-Hyacinthe (n = 401) respectively in Ontario and Quebec, Canada. Baseline data were collected in 2012 with follow-ups in 2014 and 2016. In Latin American and Albanian sites, participants were recruited randomly from their neighborhood primary care centers. In Canadian sites, letters were sent by family physicians to appropriate patients inviting them to contact the field coordinator for more information about the study. This was done to comply with the ethics guidelines at Queen's University and the University of Montreal. Approximately 30 per cent of those invited to participate in Canadian sites contacted the field coordinator to get information about the study. Of those, 95 per cent enrolled.

In Latin American and Albanian sites, the response rates were almost 100 per cent. We expected that the response rate in Canadian sites would not be as high when compared to Latin American and Albanian sites, as suggested by existing literature (Galea & Tracy, 2007; Helliwell, Aylesworth, McDowell, Baumgarten, & Sykes, 2001; Wong, Pelaez, Palloni, & Markides, 2006). In Saint-Hyacinthe, the sample of participants was representative of the community's older adult population as documented in the 2006 Canadian Census in terms of marital status, education, and income. In Kingston, participants are more highly educated than the community norm but are representative in terms of income and marital status (again, according to Canadian census data for Kingstonians, ages 65 to 74). Data collection was carried out at all sites through structured face-to-face interviews at the participants' homes unless they requested otherwise.

The IMIAS questionnaire comprised detailed structured interviews and included a wide range of measures of personal and social circumstances, self-report of existing medical conditions, health behaviours, quality of life, physical development and functioning, along with assessment of grip strength, vision, and blood pressure. Standardized training was offered to all interviewers

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

at each site. All procedures, including data collection documents and manuals were available in local site languages. Potential participants were excluded if they had four or more errors in the Orientation Scale of the Leganes Cognitive Test (LGT) (De Yebenes et al., 2003) which was administered on initial contact. Low LCT scores were considered indicative of an inability to meet the study requirements. Five people were excluded in Natal for this reason; exclusions numbered three people in Manizales and Tirana, and zero in Kingston and Saint-Hyacinthe.

Measures of Study

Since our questions on gender roles were only a part of a large project of mobility in aging, we privileged the use of a shorter version of the BSRI in older adults. For the purpose of this study, gender-stereotyped traits were measured using the 12-item BSRI short form initially used in the Spanish population (Mateo & Fernández, 1991). The masculinity scale consists of six traits that are traditionally related to instrumental behaviours and attitudes that are characterized, and perhaps stereotyped, as masculine. These traits are as follows: acting as a leader, being dominant, having leadership abilities, having a strong personality, defending one's own beliefs, and making decisions easily. The femininity scale consists of six traits that are traditionally related to expressive behaviours and attitudes that are characterized/stereotyped as feminine, specifically: being gentle, warm, sympathetic, tender, affectionate, and sensitive to others' needs. Participants were asked to rate the extent to which these items describe themselves using a 7-point Likert scale. This part of the IMIAS questionnaire took approximately 10 minutes to complete. The short form of BSRI used in this study was translated into French, Portuguese, and Albanian languages by two bilingual researchers from each IMIAS site. All the translated versions were then back-translated into English by a native English speaker to ensure accuracy and conceptual equivalency of items (Brislin, Thorndike, & Lonner, 1973) (Table 1). These measures were done to ensure that the BSRI questions had the same general meanings in all IMIAS sites.

Statistical Analysis

Preliminary statistical analysis was performed using IBM SPSS software, version 21, and STATA software, version 11. Participants with missing values on any item were excluded from the current analysis (n = 45). The majority of those participants were from Manizales, Colombia (53.3%). They were not different from those included in the analysis in terms of age, sex, years of schooling, education level, occupation type, and income sufficiency. Table 2 shows the socio-demographic characteristics of participants included in the study.

For the purpose of this article, we randomly split the combined IMIAS population samples into two data sets of approximately equal size: a "training sample" (n = 971) and a "holdout sample" (n = 979). EFA was conducted on the training sample. We performed *principal component analysis* (PCA), which is of an exploratory nature, to examine the link between the observed items and the latent constructs and to identify the factor

Table 1: English items of the 12-item short form of BSRI used in this study along with French, Spanish, Portuguese, and Albanian translations

	English	French	Spanish	Portuguese	Albanian
Femininity (expressiveness)	Gentle	Doux(ce)	Gentil	Educado	Xhentil
	Sympathetic	Sympathique	Simpatico	Compreensivo	I kuptueshem me shqetesimet e te tjereve
	Tender	Tendre	Tierno	Gentil	I bute
	Warm	Chalereux(se)	Calido	Entusiasmado	I ngrohet
	Affectionate	Affectueux(se)	Afectuoso	Carinhoso	I dashur
	Sensitive to other needs	Attentif(ve) aux besoins des autres	Sensible a lasnecesidades de los demás	Sensivel as necessidades dos outros	I ndjeshem ne nevojat e te tjereve
Masculinity	Has leadership ability	Avoir des qualités de chef (leader)	Con Madera de lider	Capacidade de liderança	Me aftesi drejtuese
(instrumentality)	Acts as a leader	Se comporter comme un leader	Actua como lider	Age como líder	Sillet si lider
	Dominant	Dominant(e)	Dominante	Dominante	Dominues
	Strong personality	Personnalité forte	Personalidad fuerte	Personalidade Forte	Me personalitet te forte
	Defends own beliefs	Deféndre ses croyances	Defensor de las propias ideas	Defende as próprias crenças	Mbron idete e veta
	Makes decisions easily	Prendre facilement des décisions	Toma decisiones facilmente	Toma decisões facilmente	Merr vendime lehte.

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

Table 2: Distribution of study participants according to demographic and socioeconomic characteristics ($n = 1,950$)

Variables	Natal (Brazil)		Manizales (Colombia)		Tirana (Albania)		Saint Hyacinthe (Quebec)		Kingston (Ontario)	
	Men ($n = 192$)	Women ($n = 210$)	Men ($n = 187$)	Women ($n = 189$)	Men ($n = 183$)	Women ($n = 204$)	Men ($n = 188$)	Women ($n = 204$)	Men ($n = 184$)	Women ($n = 209$)
Age		**$p = 0.591$***		**$p = 0.252$**		**$p = 0.279$**		**$p = 0.338$**		**$p = 0.183$**
65–69	52.1	54.8	56.7	50.8	46.40	52.0	66.0	61.8	59.8	53.1
70–74	47.9	45.2	43.3	49.2	53.60	48.0	34.0	38.2	40.2	46.9
Education		**$p = 0.002$**		**$p = 0.000$**		**$p = 0.000$**		**$p = 0.299$**		**$p = 0.265$**
Less than secondary	70.3	84.8	68.4	77.8	7.7	14.2	7.4	6.9	1.1	0.0
Secondary	22.9	12.9	13.4	17.5	19.7	38.8	38.8	46.6	23.4	21.1
Post-secondary	6.8	2.4	18.2	4.8	72.7	52.9	53.7	46.6	75.5	78.9
Income sufficiency		**$p = 0.246$**		**$p = 0.737$**		**$p = 0.017$**		**$p = 0.004$**		**$p = 0.226$**
Very sufficient	4.7	3.3	5.9	4.2	42.6	32.8	52.7	37.3	62.5	60.3
Sufficient	25.0	19.0	25.1	24.3	42.1	40.7	42.6	52.5	34.2	32.5
Insufficient	70.3	77.6	69.0	71.4	15.3	26.5	4.8	10.3	3.3	7.2
Living arrangements		**$p = 0.014$**		**$p = 0.020$**		**$p = 0.000$**		**$p = 0.000$**		**$p = 0.000$**
Alone	5.2	7.6	15.0	13.2	3.3	16.2	14.4	36.8	19.6	41.6
Only spouse	25.0	13.8	23.0	12.7	50.8	38.2	74.5	57.8	43.5	46.4
Children with or w/o spouse	69.8	78.6	62.0	74.1	45.9	45.6	11.2	5.4	37.0	12.0
Longest held occupation		**$p = 0.002$**		**$p = 0.000$**		**$p = 0.118$**		**$p = 0.372$**		**$p = 0.049$**
Not Skilled manual	35.9	52.9	29.9	50.8	8.2	6.9	23.4	26.0	15.8	12.4
Skilled manual	53.6	36.7	50.3	36.5	61.2	52.5	29.8	23.5	13.0	6.7
Non-manual	10.4	10.5	19.8	12.7	30.6	40.7	46.8	50.5	71.2	80.9
Work situation (Last week)		**$p = 0.000$**		**$p = 0.000$**		**$p = 0.472$**		**$p = 0.122$**		**$p = 0.133$**
Worked	22.4	44.8	30.5	56.1	5.5	3.9	22.9	16.7	19.0	25.4
Don't work	77.6	55.2	69.5	43.9	94.5	96.1	77.1	83.8	81.0	74.6

* ***p* values comparing men and women at each research site.**

structure (Tabachnick, 2013). The inter-relationships between the BSRI inventory dimensions were assessed using Pearson correlation coefficients. Identification of the potential number of factors was informed by the Kaiser Guttman rule of eigenvalues > 1, Cattell's scree plots, and parallel analysis (Cattell, 1966; Hayton, Allen, & Scarpello, 2004; Kaiser, 1960). The internal consistency of BSRI dimensions was assessed using Cronbach's alpha reliability coefficient. Alpha values higher than 0.7 are considered acceptable (Tavakol & Dennick, 2011). Known group validity, which is the ability of the short-form BSRI to distinguish participants of one group from another, was evaluated by comparing men and women using an independent *t*-test.

The resulting model from EFA was confirmed on the holdout sample using CFA. We then compared this model with the original two-factor structure proposed by Bem. CFA was what we used to assess measurement models with respect to goodness-of-fit indices as well as convergent and discriminant validity. Large and statistically significant factor loadings indicate that items associated with the same latent variables are highly inter-correlated, supporting convergent validity (Brown, 2006; Kline, 2011; O'Leary-Kelly & Vokurka, 1998). To assess discriminant validity, we used chi-square difference testing to compare two models, one in which latent variables are correlated and one in which they are not. Statistically significant differences in the chi-square test between the two models support discriminant validity (Segars, 1997; Zait & Bertea, 2011). For these measurement models, we used structural equation modeling (SEM) via IBM SPSS AMOS 19 software to assess relationships between factors and their observed measures. First, we assessed the multivariate normality in the holdout data set to determine the appropriate CFA estimation method. Second, we employed the modification indices to measure the amount by which the overall model chi-square test would be reduced if a parameter that was previously fixed to zero was then estimated freely (Silvia & MacCallum, 1988). Finally, we tested the fit of the models in the holdout sample, and for men and women.

Assessing the Degree of Model Fit

We used the AMOS 19 software to assess the global fit of the models using fit statistics that possess different computational logic (Hoyle, 2000). Six goodness-of-fit statistics were calculated which included chi-square, goodness-of-fit index (GFI), Bentler comparative fit index (CFI), root mean square error of approximation (RMSEA), Akaike's information criteria (AIC), and the Browne-Cudeck criterion (BCC). It is generally accepted that a statistically non-significant chi-square test indicates a good model fit, while a statistically significant chi-square test suggests that the model has a poor fit to data. However, the chi-square test is sensitive to sample size, being almost always statistically significant in larger samples such as ours, which might erroneously suggest a poor fit (Schreiber, Nora, Stage, Barlow, & King, 2006).

Therefore, other goodness-of-fit indices are recommended in addition to the chi-square test. GFI values close to 1 are indicative of good fit (Schreiber et al., 2006). Values for CFI around 0.90 are considered acceptable, whereas values around 0.95 suggest a good model fit (Bentler & Bonett, 1980; Hu & Bentler, 1999). RMSEA values below 0.05 indicate a good fit with values between 0.05 and 0.08 indicating a reasonable fit, those between 0.08 and 0.1 suggesting only a mediocre fit, and, finally, values greater than 0.10 indicating a poor model fit (Diamantopoulos & Siguaw, 2000; Hu & Bentler, 1999; Kahn, 2006; Steiger, 2007). We compared the model resulting from EFA and the two-factor structure model of Bem. The AIC and the BCC were used as measures of the final model's likelihood of being replicated in other samples of a similar size and population (Byrne, 2001; Hu & Bentler, 1999). The model with the lowest values of both AIC and BCC is more likely to yield a good fit in other samples.

Results

Principal Component Analysis

As preliminary steps for the PCA, we carried out the Kaiser Meyer-Olkin (KMO) test of sampling adequacy and Bartlett's test of sphericity on the training sample. The KMO index (0.745) is greater than 0.60–0.70 which means the data set is adequate for analysing EFA results (Netemeyer, O'Bearden, & Sharma, 2003). Bartlett's test rejects the hypothesis that the correlation matrix is an identity matrix, without significant correlations between variables (at $p < 0.000$) (Tabachnick & Fidell, 2001). Both tests confirm that the training sample is suitable for factor analysis.

Based on the guidance of the Kaiser-Guttman rule, scree plot, and parallel analysis test (Figure 1), we determined that a three-factor structure was most appropriate.

Oblique rotations using the Promax method (Kappa = 4) were generated to assist in factor structure interpretation because we found that factors 1 and 3, and factors 2 and 3, were slightly correlated (see Table 3) (Hendrickson & White, 1964).

Inspection of the pattern matrix revealed a three-factor structure. All of the 12 items had factor loadings greater than 0.45 following the criteria of Comrey and Lee (1992), communalities were greater than or equal to 0.43, and no items failed to show salient loadings on any of the factors. The three factors accounted for 56.8 per cent of variance in scores. Factor 1 (femininity or

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

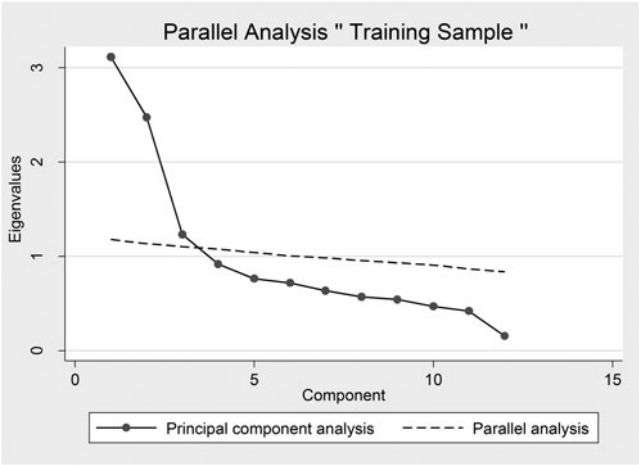

Figure 1: Difference in eigenvalues from parallel analysis

expressiveness factor) accounted for 25.94 per cent of the variance (eigenvalue = 3.11). Factor 2 (masculinity or instrumentality factor) accounted for 20.60 per cent of the variance (eigenvalue = 2.47). Factor 3 (mixed factor) accounted for 10.26 per cent of the variance (eigenvalue = 1.23). The expressiveness factor had five items, the instrumentality factor had four items, and the mixed factor had three items. Factor loadings, communalities, and mean score and standard deviation of each item are shown in Table 4.

Confirmatory Factor Analysis

A CFA framework was applied to the whole holdout sample, also to the aggregate holdout sample of men and women. The holdout sample showed multivariate non-normal distribution with a value of more than five for Mardia's coefficient. We used the *asymptomatic distribution-free* (ADF) estimation which is a trustworthy estimation method recommended for CFAs under non-normal conditions, provided that the sample size is adequate (Raykov & Marcoulides, 2006). Before doing CFA, we calculated internal consistency for the three factors. Cronbach's alphas were 0.77, 0.78, and 0.47 for expressiveness, instrumentality, and the mixed factor respectively. The first two factors showed acceptable reliability whereas for the third factor reliability was poor (Tavakol & Dennick, 2011). We also calculated the reliability coefficients for the two factors structure as proposed by Bem, and found alphas values of 0.76 for the expressiveness factor and 0.75 for the instrumentality factor. These values indicate acceptable alphas for the two-factor model.

Table 3: Correlation matrix of the three factor structure of the 12-item short-form BSRI in training sample ($n = 971$)

Dimension	Factor 1	Factor 2	Factor 3
Factor 1		0.01	0.16*
Factor 2	0.01		0.25*
Factor 3	0.16*	0.25*	

The factor label components are as follows: F1, femininity (expressiveness) factor; F2, masculinity (instrumental) factor, F3 mixed factor
*** $p < 0.05$.**

Next, the three-factor model consistent with factor structure from the EFA (model 1), along with the two-factor model (model 2) proposed by Bem, were considered using CFA to assess how well data fit both models. One common assumption in these models is that a parameter is fixed at zero (e.g., in a two-factor model, items related to one factor are free to load on it while restricted to have zero loadings on the other factor). Modification indices were employed to assess how reasonable these assumptions were when they were relaxed (Silvia & MacCallum, 1988). The results indicated that a two-factor structure (model 2) had a satisfactory fit, and it clearly demonstrated better fit than a three-factor solution (model 1). Specifically, the fit indices for model 2 in the whole holdout sample are as follows: $\chi^2 = 125.34$, GFI = 0.96, CFI = 0.93, and RMSEA = 0.046. The same goodness-of-fit indices for model 1 are as follows, in sequence: 218.99, 0.93, 0.86, 0.06.

We obtained similar results when fitting models for men and women separately (see Table 5). Model 2 showed that all un-standardized factor loadings were statistically significant at $p < 0.001$, rejecting the null hypothesis that the unconstrained loadings are zero. All standardized factor loadings showed moderate to high values, indicating that each item is at least moderately related to its corresponding latent variable (Figure 2). Only item 10 ("defends own beliefs") showed low standardized factor loading (0.22). We assessed a model without this item, but the goodness-of-fit indices were worse compared to the 12-item, two factor model. Therefore, we consider model 2 with acceptable convergent validity. Using the entire IMIAS data set, we compared CFA results of model 2 in which the two latent variables correlate, with the same model in which they did not correlate. The chi-square difference test ($\chi^2 - \chi^2 = 32.82$; $df_1 \_ df_2 = 1$) was statistically significant ($p < 0.001$), indicating discriminant validity.

Known Group Validity

Known group validity was used to determine the extent to which the two-factor structure of the 12-item BSRI revealed known differences in gender role stereotypes between men and women. In the whole sample, we found statistically significant differences between men and women on both dimensions using independent samples *t*-tests. In general, higher mean values

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

Table 4: Principal component analysis of the 12-item short-form BSRI with communalities (h^2) of each item in training sample ($n = 971$)

No.	Item	Factor 1	Factor 2	Factor 3	h^2	*Mean*	*SD*
1	Gentle	0.68	0.00	0.00	0.48	5.86	1.30
2	Sympathetic	0.72	0.00	0.00	0.53	6.03	1.12
6	Tender	0.79	0.00	0.00	0.60	5.53	1.44
7	Warm	0.59	0.00	0.00	0.43	5.38	1.65
8	Affectionate	0.78	0.00	0.00	0.61	5.61	1.51
3	Has leadership abilities	0.00	0.92	0.00	0.82	4.38	2.01
4	Act as a leader	0.00	0.92	0.00	0.81	4.09	1.96
5	Dominant	0.00	0.58	0.00	0.48	3.61	1.88
9	Strong personality	0.00	0.48	0.00	0.53	4.94	1.77
10	Defends own beliefs	0.00	0.00	0.78	0.60	5.79	1.40
11	Sensitive to others needs	0.00	0.00	0.51	0.45	5.99	1.22
12	Make decisions easily	0.00	0.00	0.63	0.48	4.96	1.69
	% of variance	**25.94**	**20.60**	**10.26**			
	Alpha	**0.77**	**0.78**	**0.47**			

The factor label components are as follows: F1, femininity (expressiveness) factor; F2, masculinity (instrumental) factor; F3, mixed factor
The factor pattern coefficients of 0.45 and below were replaced by zeros.
Items were scored (Never/almost never true = 1, almost/always true = 7).

were found for expressiveness among women and instrumentality for men in the aggregate sample and when data were disaggregated by site (see Table 6). Mean masculinity scores were higher in men compared to women in Tirana, St-Hyacinthe, and Kingston ($p < 0.05$), while mean femininity scores were higher in women than men in Manizales, Tirana, and Kingston ($p < 0.05$).

Discussion

According to previous literature, the original and the abbreviated versions of the BSRI have both been used primarily among younger populations, and knowledge about validity in older-adult populations is limited. We decided to use the Bem Sex Role Inventory as a measure of gender-stereotyped traits because of recent evidence on the salience of these traditional

Table 5: Goodness-of-fit indices for the three-factor and the two-factor models of the 12-item short-form BSRI in the holdout sample of the IMIAS population

Model 1	**Three-factor Model**							
	Number of Parameters = 30							
	χ^2	***df***	***p***	**GFI**	**CFI**	**RMSEA**	**AIC**	**BCC**
Total sample (n = 979)	218.99	48	0.00	0.93	0.86	0.06	278.98	279.79
Men (n = 462)	112.49	48	0.00	0.93	0.89	0.054	172.49	174.23
Women (n = 517)	155.11	48	0.00	0.93	0.86	0.066	215.11	216.66
Model 2	**Two-factor Model**							
	Number of Parameters = 37							
	χ^2	***df***	***p***	**GFI**	**CFI**	**RMSEA**	**AIC**	**BCC**
Total sample (n = 979)	125.34	41	0.00	0.96	0.93	0.046	199.34	200.33
Men (n = 462)	68.48	41	0.01	0.96	0.95	0.038	142.48	144.63
Women (n = 517)	107.54	41	0.00	0.95	0.91	0.056	181.54	183.46

AIC = Akaike's information criteria
BCC = Browne-Cudeck criterion
CFI = comparative fit index
GFI = goodness-of-fit index
IMIAS = International Mobility in Aging Study
RMSEA = root mean square error of approximation

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

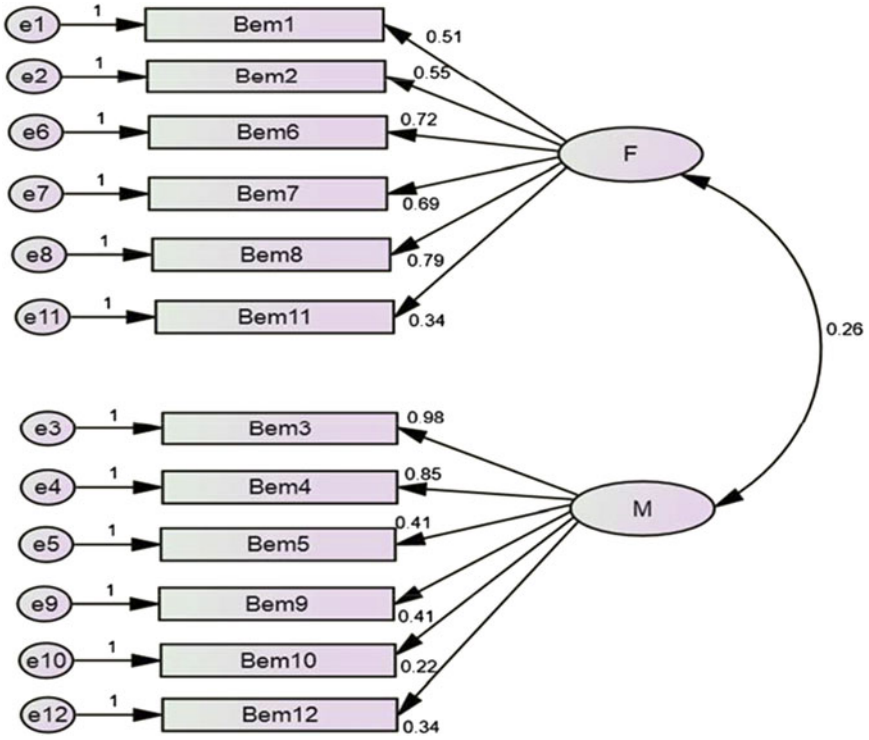

Figure 2: Hypothesised two-factor model (model 2) of factorial structure of Bem Sex Role Inventory ($n = 979$). Bem 1 = Gentle; Bem 2 = Sympathetic; Bem 3 = Has leadership abilities; Bem 4 = Act as a leader; Bem 5 = Dominant; Bem 6 = Tender; Bem 7 = Warm; Bem 8 = Affectionate; Bem 9 = Strong personality; Bem 10 = Defends own belief; Bem 11 = Sensitive to others' needs; Bem 12 = Make decisions easily. Standardized factor loadings appear on the lines. All loadings are statistically significant at $p < 0.00$

gender roles in Latin America and Southern Europe (Chant & Craske, 2003; Miluka, 2009; Särnhult, 2014; Silova & Magno, 2004; Stecklov, Carletto, Azzarri, & Davis, 2010). In addition, research findings obtained from a sample of older adults from Canada refer to these same gender roles as personality attributes having significant effects on general wellness and life satisfaction in older women (Gale-Ross, Baird, & Towson, 2009). We hypothesized that gender roles may be more applicable to older generations than to younger populations who have more opportunity to choose or to explore more flexible gender roles, and this hypothesis needs to be investigated. The purpose of this analysis was to evaluate the psychometric properties and construct validity of the 12-item short-form BSRI in older adults and in developing as well as developed countries.

We conducted an exploratory and then a confirmatory factor analysis of the 12-item BSRI. The exploratory factor analysis revealed a three-factor model. This model was further confirmed by CFA and compared with the two-factor model initially proposed by Bem. The CFA revealed that the two-factor solution showed satisfactory construct validity and superior fit compared with the three-factor solution, according to the standard recommendations of SEM literature (Diamantopoulos & Siguaw, 2000; Hu & Bentler, 1999; Kahn, 2006; Steiger, 2007). In addition, the two-factor model showed the lowest values of both AIC and BCC, indicating that it is more likely to reproduce findings in other samples of similar populations. Discrepancy between results from exploratory and confirmative factor analysis is not surprising because EFA is data driven and involves a number of subjective decisions (Brown, 2006). In terms of reliability, the CFA two-factor solution showed higher than the recommended cut-off point of 0.70 (Bland & Altman, 1997; Tavakol & Dennick, 2011), indicating acceptable internal consistency for the whole IMIAS sample ($\alpha = 0.75$ and $\alpha = 0.76$ for instrumentality and expressiveness items respectively) and for men and women separately (α ranged between 0.73 and 0.78). A significant gender difference was found in the two factors. Men tended to have significantly higher instrumentality and lower expressiveness scores than women. These findings confirmed expected differences in gender stereotype compositions assumed by men and women.

There are several justifications for the existence of two factors as proposed by Bem, even where more than two factors exist. Gender role orientations may vary across cultural settings or due to changes in self-concepts of gender roles over time and in response to economic,

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

Table 6: Comparison between men ($n = 462$) and women ($n = 517$) responses on two-factor model of the 12-item short form of the Bem Sex Role Inventory in the confirmatory holdout sample of the IMIAS population

	Men	Women	
Dimension	**Mean (*SD*)**	**Mean (*SD*)**	***t***
All samples			
Femininity (expressiveness)	5.62 (0.96)	5.88 (0.88)	–4.43*
Masculinity (instrumentality)	4.77 (1.13)	4.49 (1.20)	3.68*
Natal (Brazil)			
Femininity (expressiveness)	5.29 (1.04)	5.34 (0.92)	–0.31
Masculinity (instrumentality)	4.36 (1.3)	4.18 (1.3)	0.93
Manizales (Colombia)			
Femininity (expressiveness)	5.62 (1.17)	6.09 (0.99)	–2.92*
Masculinity (instrumentality)	4.32 (1.1)	4.28 (1.1)	0.22
Tirana (Albania)			
Femininity (expressiveness)	6.03 (0.77)	6.35 (0.62)	–3.19*
Masculinity (instrumentality)	5.07 (1.05)	4.43 (1.1)	4.13*
Saint Hyacinthe (Quebec)			
Femininity (expressiveness)	5.72 (0.77)	5.78 (0.72)	–0.58
Masculinity (instrumentality)	4.89 (1.01)	4.39 (1.14)	3.27*
Kingston (Ontario)			
Femininity (expressiveness)	5.41 (0.80)	5.83 (0.82)	–3.63*
Masculinity (instrumentality)	5.22 (0.82)	5.09 (1.05)	0.96

*** $p < 0.05$**
IMIAS = International Mobility in Aging Study

social, and political transitions. For instance, research findings from studies in the United States and Latin America suggest that stereotypes of women and women's self-perceptions have become more instrumental in line with their increasing emancipation and participation in public life (Diekman & Eagly, 2000; Diekman, Eagly, Mladinic, & Ferreira, 2005). Again, the BSRI was developed to assess the extent to which traditional cultural norms of desirable female and male attributes reflect individual self-perceptions (Bem, 1979). It could be argued that the North American norms of the 1970s, upon which the BSRI was constructed, may define concepts of masculinity and femininity that are no longer applicable in North America and whose applicability elsewhere was never established. However, in this work we provided evidence for all four gender roles in both Canadian cities and also in very different international contexts supporting the ubiquitous existence of these four types. Other researchers suggest that the factor structure of BSRI may be more complex than Bem's proposal of masculine or feminine items split into two factors (Feather, 1978; Maznah & Choo, 1986; Ratliff & Conley, 1981). This means that it may be inappropriate to define masculinity and femininity based on single factors when more than one factor might be embedded in each concept. Despite these concerns, our findings support Bem's two-factor model and are consistent with most of the recent literature on BSRI validation using CFA in a sample where older adults were over-represented (Sellars, 2008).

One of the major strengths of this study is a large sample size (n = 1,950). Furthermore, we applied EFA to a sample of 971 participants; a high ratio of participants to BSRI items (ratio 81:1) ensured that stable factors could be identified using EFA (Gorsuch, 1983). CFA was conducted with a separate, large sample (n = 979). Our findings therefore provide strong support for the 12-item short form BSRI as a valid measure of gender roles as reflected by individual sex roles among older adults.

The results of this study strongly support the psychometric properties and construct validity of the 12-item short-form BSRI among an international sample of older adults. However, limitations of the study must be acknowledged. First, in less-educated participants, BSRI may contribute to measurement errors and information bias. We attempted to minimize these errors by administering the BSRI at all sites using responsive visual aids. Second, the Kingston sample was overeducated compared with the 2006 Kingston population census data for this 65–74-year-old age group. Consequently, participants included in the study from Kingston may not be representative of their community, and generalization of results from this specific site may be of limited validity. However, the diverse samples provide a wide distribution range of social variables for analysis, reinforcing the validity of the instrument. Finally, the instrument is a self-reported measure that allows participants to rate themselves in some aspects

https://doi.org/10.1017/S0714980816000404 Published online by Cambridge University Press

of common cultural values. We cannot exclude possible social desirability bias because we cannot determine the extent to which responses accurately reflect participants' actual characteristics, behaviours, and experiences.

Conclusions

The 12-item short form BSRI appears to be a brief, robust, reliable, and psychometrically sound instrument that can be readily implemented in studies to assess gender roles. It offers a method for studying whether gender norms, expectations, and constraints have an impact on outcomes such as mental and physical health. Further use of the current 12-item BSRI is justified and will provide additional evidence to refine and compare its psychometric properties and construct validity among older-adult populations.

Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/S0714980816000404

References

Annandale, E., & Hunt, K. (1990). Masculinity, femininity and sex: An exploration of their relative contribution to explaining gender differences in health. *Sociology of Health Illness*, *12*(1), 24–46.

Ballardreisch, D., & Elton, M. (1992). Gender orientation and the Bem Sex-Role Inventory: A psychological construct revisited. *Sex Roles*, *27*(5–6), 291–306.

Bazik, N. (2011). *Gender role attitudes in youth* (unpublished doctoral dissertation). American University, Washington, DC.

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

Bem, S. L. (1979). Theory and measurement of androgyny: Reply. *Journal of Personality and Social Psychology*, *37*(6), 1047–1054.

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, *314*, 572.

Bollen, K. (1989). *Structural equations with latent variables*. Chapel Hill, NC: Wiley.

Brislin, R. W., Thorndike, R. M., & Lonner, W. J. (1973). *Cross-cultural research methods*. New York, NY: Wiley.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Burke, P. J., Stets, J. E., & Piroggood, M. A. (1988). Gender identity, self-esteem, and physical and sexual abuse in dating relationships. *Social Psychology Quarterly*, *51*(3), 272–285.

Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, *1*(1), 55–86.

Carver, L. F., Vafaei, A., Guerra, R., Freire, A., & Phillips, S. P. (2013). Gender differences: Examination of the 12-item Bem Sex Role Inventory (BSRI-12) in an older Brazilian population. *Plos One*, *8*(10), e76356.

Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Chant, S. H., & Craske, N. (2003). *Gender in Latin America*. London, UK: Rutgers University Press.

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*(5), 872–887.

Choi, N., Fuqua, D. R., & Newman, J. L. (2009). Exploratory and confirmatory studies of the structure of the Bem Sex Role Inventory short form with two divergent samples. *Educational and Psychological Measurement*, *69*(4), 696–705.

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*(3), 384–387.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Los Angeles, CA: Taylor & Francis.

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

De Yebenes, M. J., Otero, A., Zunzunegui, M. V., Rodriguez-Laso, A., Sanchez-Sanchez, F., & Del Ser, T. (2003). Validation of a short cognitive tool for the screening of dementia in elderly people with low educational level. *International Journal of Geriatric Psychiatry*, *18*(10), 925–936.

Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. Vienna, AUT: Sage.

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*(10), 1171–1188.

Diekman, A. B., Eagly, A. H., Mladinic, A., & Ferreira, M. C. (2005). Dynamic stereotypes about women and men in Latin America and the United States. *Journal of Cross-Cultural Psychology*, *36*(2), 209–226.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558.

Feather, N. T. (1978). Factor structure of the Bem Sex-Role Inventory: Implications for the study of masculinity, femininity, and androgyny. *Australian Journal of Psychology*, *30*(3), 241–254.

Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of Epidemiology*, *17*(9), 643–653.

Gale-Ross, R., Baird, A., & Towson, S. (2009). Gender role, life satisfaction, and wellness: Androgyny in a southwestern Ontario sample. *Canadian Journal on Aging*, *28*(2), 135–146.

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gough, H. G. (1952). Identifying psychological femininity. *Educational and Psychological Measurement*, *12*(3), 427–439.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205.

Helliwell, B., Aylesworth, R., McDowell, I., Baumgarten, M., & Sykes, E. (2001). Correlates of nonparticipation in the Canadian Study of Health and Aging. *International Psychogeriatrics*, *13*, 49–56.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*(1), 65–70.

Hoffman, R. M., & Borders, L. D. (2001). Twenty-five years after the Bem Sex-Role Inventory: A reassessment and new issues regarding classification variability. *Measurement and Evaluation in Counseling and Development*, *34*(1), 39–55.

Holt, C. L. (1998). Assessing the current validity of the Bem Sex-Role Inventory. *Sex Roles*, *39*(11–12), 929–941.

Hoyle, R. H. (2000). 16-Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465–497). San Diego, CA: Academic Press.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist*, *34*(5), 684–718.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Katsurada, E., & Sugihara, Y. (1999). A preliminary validation of the Bem sex role inventory in Japanese culture. *Journal of Cross-Cultural Psychology*, *30*(5), 641–645.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Lindsey, L. L. (2005). *Gender roles: A sociological perspective*. Maryville, MO: Pearson Prentice Hall.

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*(4), 427–449.

Mateo, M., & 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*(1), 31–41.

Miluka, J. (2009). *Education, migration, and labor markets in Albania: A gender perspective*. Washington, DC: ProQuest.

Netemeyer, R. G., O'Bearden, W., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387–405.

Özkan, T., & Lajunen, T. (2005). Masculinity, femininity, and the Bem Sex Role Inventory in Turkey. *Sex Roles*, *52*(1–2), 103–110.

Perry, D. G., & Bussey, K. (1979). The social learning theory of sex differences: Imitation is alive and well. *Journal of Personality and Social Psychology*, *37*(10), 1699–1712.

Ratliff, E. S., & Conley, J. (1981). The structure of masculinity-femininity: Multidimensionality and gender differences. *Social Behavior and Personality*, *9*(1), 41–47.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. East Lansing, MI: Lawrence Erlbaum Associates.

Särnhult, V. (2014). *Gender and power-images of female politicians in Colombia: A critical discourse analysis* (unpublished doctoral dissertation). Stockholm University, Sweden.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338.

Segars, A. H. (1997). Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research. *Omega-International Journal of Management Science*, *25*(1), 107–121.

Sellars, B. C. (2008). *A longitudinal view of sex role development: Demographic differences and the influence of social relations*. Ann Arbor, MI: ProQuest.

Silova, I., & Magno, C. (2004). Gender equity unmasked: Democracy, gender, and education in Central/Southeastern Europe and the former Soviet Union. *Comparative Education Review*, *48*(4), 417–442.

Silvia, E. S. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behavioral Research*, *23*(3), 297–326.

Spence, J. T. (1984). Masculinity, femininity, and gender-related traits: A conceptual analysis and critique of current research. *Progress in Experimental Personality Research*, *13*, 1–97.

Spence, J. T. (1985). Gender identity and its implications for the concepts of masculinity and femininity. *Nebraska Symposium on Motivation*, *32*, 59–95.

Stecklov, G. U. Y., Carletto, C., Azzarri, C., & Davis, B. (2010). Gender and migration from Albania. *Demography*, *47*(4), 935–961.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–898.

Stets, J. E., & Burke, P. J. (2000). *Femininity/masculinity*. New York, NY: Macmillan

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, MA: Pearson.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.

Vafaei, A., Alvarado, B., Tomás, C., Muro, C., Martinez, B., & Zunzunegui, M. V. (2014). The validity of the 12-item Bem Sex Role Inventory in older Spanish population. An examination of the androgyny model. *Archives of Gerontology and Geriatrics*, *59*(2), 257–263.

Williams, J. E., & Best, D. L. (1982). *Measuring sex stereotypes: A thirty-nation study*. Winston-Salem, NC: Sage.

Williams, J. E., & Best, D. L. (1990). *Sex and psyche: Gender and self viewed cross-culturally*. Winston-Salem, NC: Sage.

Windle, M., & Sinnot, J. D. (1985). A psychometric study of the Bem Sex-Role Inventory with an older adult sample. *Journals of Gerontology*, *40*(3), 336–343.

Wong, R., Pelaez, M., Palloni, A., & Markides, K. (2006). Survey data for the study of aging in Latin America and the Caribbean: Selected studies. *Journal of Aging and Health*, *18*(2), 157–179.

Zait, A., & Bertea, P. E. (2011). Methods for testing discriminant validity. *Management & Marketing Journal*, *9*(2), 217–224.

Zhang, J., Norvilitis, J. M., & Jin, S. H. (2001). Measuring gender orientation with the Bem Sex Role Inventory in Chinese culture. *Sex Roles*, *44*(3–4), 237–251.