JOURNAL OF PERSONALITY ASSESSMENT, *84*(3), 271–278

Cross-Cultural Validity of the Masculine and Feminine Gender Role Stress Scales

Sonja van Well and Annemarie M. Kolk
Department of Clinical Psychology
University of Amsterdam

Willem A. Arrindell
Department of Clinical Psychology
University of Groningen

The objective was to examine the usefulness of Dutch versions of the Masculine Gender Role Stress (MGRS; Eisler & Skidmore, 1987) Scale and the Feminine Gender Role Stress (Gillespie & Eisler, 1992) Scale in The Netherlands. Undergraduate students ($N = 2,239$) completed both gender role stress scales. A subgroup ($n = 508$) also completed questionnaires about masculinity–femininity and daily hassles. With regard to both gender role stress scales, results of confirmatory factor analyses supported the original 5-factor structures and revealed no cross-sex differences on the factor models. Reliability and homogeneity indexes were all well within acceptable to satisfactory limits. Further evidence of construct validity was found in (a) medium to large correlations with daily hassles, (b) sex differences on the FGRS scale, and (c) small to medium correlations with masculinity–femininity. The major discrepancy with previous studies was that for Dutch female and male students, the MGRS scale was not sex specific. Taken together, this study sustained the utility of both gender role stress scales for use in The Netherlands.

To investigate the relation between masculinity–femininity and health, early gender research used instruments such as the Bem Sex Role Inventory (BSRI; Bem, 1974) and the Personal Attributes Questionnaire (PAQ; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974). These instruments assess masculinity–femininity by means of self-ascribed personality characteristics. The BSRI contains personality characteristics more desirable for one sex than for the other, whereas the PAQ comprises personality characteristics that are equally desirable for men and women but stereotypically more strongly so for one sex than for the other. Although associations between masculinity–femininity and health have been found, research on masculinity–femininity as measured by the BSRI or PAQ is limited for two reasons. First, both instruments define masculinity–femininity in terms of self-ascribed personality characteristics. They fail to also take into account other components of masculinity–femininity such as attitudinal and behavioral components (McCreary, 1990). Second, because both instruments include mainly positive personality characteristics, it is difficult to find links between these instruments and health problems (Martz, Handley, & Eisler, 1995).

More recently, Eisler and colleagues (Eisler, 1995; Eisler & Blalock, 1991; Eisler & Skidmore, 1987; Gillespie & Eisler, 1992) argued that it is not the masculine or feminine behavioral or coping strategy that increases vulnerability to health problems, but it is the rigid or maladaptive gender role-determined attitude that decreases the person's repertoire of behaviors or coping skills, and this reduction increases vulnerability to health problems. They introduced the concept of gender role stress, which may better explain health problems of men and women based on gender-specific socialization. Gender role stress, as they pointed out, refers to stress resulting from one's perceived failure to meet the demands of one's adhered gender role. Rigid adherence to the culturally prescribed gender role values interferes with one's objective appraisal of a situation, which produces heightened levels of stress when facing a gender role related situation and limits the range of gender appropriate coping strategies.

To assess the degree to which certain gender role related situations are stressful, the Masculine Gender Role Stress (MGRS; Eisler & Skidmore, 1987) Scale and the Feminine Gender Role Stress (FGRS; Gillespie & Eisler, 1992) Scale were developed. The MGRS scale contains specific situa-

tions that elicit stress in relation to the perceived failure to meet the standards of the masculine gender role, for example, "appearing less athletic than a friend" and "being outperformed at work by a woman." These masculine gender role stressors involve five fears: feeling physically inadequate, expressing tender emotions, being outperformed by women in male activities, being intellectually inferior, and experiencing performance failure with regard to work and sexual activities. In contrast, the FGRS Scale consists of specific situations that elicit stress in relation to the perceived failure to live up to the feminine gender role expectations, for example, "being perceived by others as overweight" and "having someone else raise your children." The FGRS Scale contains five fears: developing unemotional relationships, being physically unattractive, being exposed to the potential harm of violence, behaving assertively, and not being nurturant.

With regard to the reliability, the MGRS subscales have high internal consistencies (with Cronbach's alpha's in the low .90s; Eisler, Skidmore, & Ward, 1988). In addition, the FGRS subscales have satisfactory internal consistencies (with Cronbach's alpha's ranging from .73 to .83), and the total FGRS scale has satisfactory test–retest reliability over a 2-week interval ($r = .82$; Gillespie & Eisler, 1992). The construct validity of both scales has also been supported. With regard to convergent validity, men scored significantly higher on the MGRS scale than women, whereas women scored significantly higher on the FGRS scale than men (Eisler & Skidmore, 1987; Gillespie & Eisler, 1992). Furthermore, the MGRS scale is related to anger ($r = .54$, $p < .01$), anxiety ($r = .23$, $p < .05$; Eisler & Skidmore, 1987; Eisler et al., 1988), and irrational fears (Arrindell, Kolk, Pickersgill, & Hageman, 1993), whereas the FGRS scale is related to daily hassles ($r = .43$, $p < .01$) and depression ($r = .31$, $p < .01$; Gillespie & Eisler, 1992). Pertaining to discriminant validity, the MGRS scale has no association with masculinity (e.g., $r = .08$, $p > .05$, and $r = –.0550$, $p > .05$), and the FGRS scale has no association with femininity ($r = .21$, $p = .06$; masculinity–femininity as measured by the PAQ). This lack of association among measures confirms the distinction between gender role stress and masculinity–femininity (Eisler & Skidmore, 1987; Eisler et al., 1988; Gillespie & Eisler, 1992).

As gender role standards are largely prescribed by cultural beliefs and expectations, it is important to address the generalization of the gender role stress construct across cultures. However, there are hardly any studies available on the cross-cultural validity of the gender role stress scales. To the best of our knowledge, there are only two studies available on the cross-cultural generality of the gender role stress concept (Tang & Lau, 1995, 1996). The first study (Tang & Lau, 1995) investigated the assessment of gender role stress in Chinese samples. Similar to Americans, Chinese men scored higher on the MGRS scale and lower on the FGRS scale than Chinese women. Furthermore, gender role stress was associated with (mental) health disturbances (Tang & Lau, 1995). The second study (Tang & Lau, 1996) addressed the factor structure of Chinese gender role stress scales. The original five-dimensional structures of both the MGRS and FGRS scales were rejected. Instead, three-factor structures were identified for both Chinese gender role stress scales. The MGRS scale includes the factors Performance Failure, Inferiority, and Emotional Inexpressiveness, whereas the FGRS scale contains the factors Inadequacy, Unassertiveness, and Victimization.

At present, we are faced with the lack of empirical data on the cross-national generalizability of both the MGRS and FGRS scales in countries other than the United States and China. Therefore, the aim of this study was to examine the usefulness of Dutch versions of the MGRS and FGRS scales in The Netherlands. Specifically, the objectives of this investigation were threefold. First, we made an attempt to study the invariance characteristics (factorial validity) of the dimensions underlying the MGRS and FGRS scales across nationality (United States vs. Dutch) and biological sex using a confirmatory approach. Second, we studied some reliability characteristics (internal consistency, homogeneity, and item-remainder associations) of the (sub)scales. Third, further evidence of construct validity was sought under the predictions that (a) men would score higher than women on the MGRS scale, whereas the opposite would be shown to hold in relation to the FGRS scale; (b) self-reports of the subjective experience of exposure to daily hassles would correlate significantly and positively with gender role stress scores; and (c) only small to negligible associations ($r < .10$; Cohen, 1992) would be demonstrated between the gender role stress scales and measures of masculinity–femininity (discriminant validity).

METHOD

Participants

Three groups of college students at the University of Amsterdam participated in the study. The first group included 360 female and 153 male undergraduate psychology students recruited from an introductory psychology course and aged between 17 and 58 years ($M = 21.5$, $SD = 5.4$). As part of a course requirement, these students completed, in five consecutive weekly sessions, a large number of computer-administered questionnaires including measures that are administered for other scientific purposes.

The second group included 831 female and 618 male undergraduate students from different disciplines (e.g., economics, law, chemistry, social science) and aged between 18 and 64 years ($M = 24.4$, $SD = 5.0$). They were approached by means of an email message, which was sent to approximately 10,000 randomly chosen students. They were requested to go to a Website containing a questionnaire about stress experience and to complete it online.

The third group consisted of 94 female and 183 male undergraduate students also from different disciplines (e.g., economics, law, chemistry, social science) and aged between 18 and 45 years ($M = 21.2$, $SD = 3.8$). They were approached during a lecture break or after a lecture and asked to complete a paper-and-pencil questionnaire about stress experience. Using this procedure, approximately 900 students were approached.

Participants of the first group received course credit, whereas the respondents of the second and third group volunteered to participate without course credit or monetary reward.

Instruments

Gender role stress. The MGRS scale (Eisler & Skidmore, 1987) was used to establish masculine gender role stress. This scale contains 40 items considered more stressful for men than for women. The MGRS scale consists of five subscales, that is, Physical Inadequacy (9 items), Emotional Inexpressiveness (7 items), Subordination to Women (9 items), Intellectual Inferiority (7 items), and Performance Failure (8 items). Feminine gender role stress was assessed by the FGRS scale (Gillespie & Eisler, 1992). This scale consists of 39 items considered more stressful for women than for men. The FGRS scale contains five subscales, that is, Unemotional Relationships (10 items), Physical Unattractiveness (8 items), Victimization (6 items), Behaving Assertively (7 items), and Not Being Nurturant (8 items). Participants rated each item on a 6-point Likert scale ranging from 0 (*not stressful*) to 5 (*extremely stressful*). Both gender role stress scales were translated from English into Dutch and cross-checked for inconsistencies through back translations to English to establish translation equivalence following standard procedures laid down in the cross-cultural psychological literature (Brislin, 1986).

Masculinity–femininity. Masculinity–femininity was first established with the BSRI (Bem, 1974; Dutch translation by Kolk & de Ruijter, 1993). This questionnaire consists of 60 personality characteristics. Of these, 20 characteristics, more desirable for men than for women, represent the Masculinity scale (e.g., independent, assertive), whereas 20 characteristics, more desirable for women than for men, make up the Femininity scale (e.g., tender, affectionate). The remaining items serve as filler items. Participants rated each characteristic as to how much it applied to them on a 7-point Likert scale ranging from 1 (*never or almost never true*) to 7 (*always or almost always true*). Both scales have been found to be reliable and valid (Bem, 1974; Holt & Ellis, 1998). Based on this data set, the following reliability figures were obtained for the Masculinity subscale: Cronbach's $\alpha = .81$, mean interitem correlation = .19, and item-remainder correlations ranging from .07 to .62. The corresponding figures for the Femininity subscale were .78, .19, and .05 to .65, respectively.

Masculinity–femininity was also measured with the PAQ (Spence & Helmreich, 1978; Spence et al., 1974; Dutch translation by van Well & Kolk, 2002). The PAQ consists of 24 trait dimensions. The Masculinity scale contains eight items stereotypically more associated with men than with women (e.g., self-confident), whereas the Femininity scale contains eight items stereotypically more associated with women than with men (e.g., understanding of others). The remaining items serve as filler items. Participants rated each item as to how much it applied to them on a 5-point scale ranging from 1 (*not at all*) to 5 (*very*). Validity and reliability of the PAQ scales were found to be satisfactory (Helmreich, Spence, & Wilhelm, 1981). Based on this data set, the following reliability figures were obtained for the masculinity subscale: Cronbach's $\alpha = .67$, mean interitem correlation = .21, and item-remainder correlations ranging from .16 to .48. The corresponding figures for the Femininity scale were .71, .25, and .30 to .53, respectively.

Daily hassles. Exposure to daily hassles was assessed by the shortened 41-item version of the Survey of Recent Life Experiences (SRLE; Kohn & Macdonald, 1992; Dutch translation by de Jong, Timmerman, & Emmelkamp, 1996). Each item describes a daily hassle such as "being let down or disappointed by friends" or "too many things to do at once." Participants rated the extent to which the items had been part of their lives over the past month on a 4-point Likert scale ranging from 1 (*not at all part of my life*) to 4 (*very much part of my life*). Kohn and MacDonald (1992) reported good internal consistency (for the short version, Cronbach's $\alpha = .90$) and construct validity. A similar reliability figure was yielded by de Jong et al. (1996) who reported a Cronbach's alpha of .89. Based on this data set, the following reliability figures were obtained: Cronbach's $\alpha = .90$, mean interitem correlation = .18, and item-remainder correlations ranging from .18 to .55.

Procedure

The MGRS and FGRS scales were combined and presented as one questionnaire set that dealt with the topic of stress experience. Completion took 10 min on average. In addition to the gender role stress scales, participants of the first group also completed the PAQ, the BSRI, and the SRLE. These measures were administered on different days and completion took about 30 min.

RESULTS

Study Sample

Data of 184 participants (12.7%) of the second group and 23 participants (8.3%) of the third group were deleted because of incomplete information (more than 30% missing values).

Thereafter, less than 0.5% missing values were observed. Corrected item mean substitutions (Huisman, 1999) were used to replace these missing values. In addition, data of 4 participants (0.8%) of the first group and 5 participants (0.4%) of the second group were deleted for representing outliers (z score > 3).

As data on the measures of masculinity–femininity and daily hassles were obtained from the first group participants only, correlations between these measures and the gender role stress scales were available for this group only. For the remaining analyses, the data of the three participant groups were pooled. The groups were similar in that they all consisted of university students (of similar age and educational level). Also the groups were comparable with regard to the previously mentioned data loss. However, the groups differed in data collection method (i.e., computer classroom vs. computer online vs. paper-and-pencil) and participation (i.e., compulsory vs. voluntary). Because no a priori hypotheses about differences between the student groups, data collection methods, or participation were posed, data of the three groups were pooled.

Factorial Validity

Intercorrelations between MGRS and FGRS subscales are presented in Table 1 and Table 2, respectively. All correlations between the subscales were significantly positive and varied from .39 to .60 for the MGRS scale and from .50 to .65 for the FGRS scale.

Confirmatory factor analyses (CFAs) were based on covariance matrices and conducted using LISREL 8 (Jöreskog & Sörbom, 1993). Fit indexes for both the MGRS and FGRS scales are presented in Table 3 for the sexes and for each factor model separately.

The fit indexes of the MGRS models were first examined for men. The absolute fit indexes standardized root mean squared residual (SRMR) and root mean squared error of approximation (RMSEA) were inspected. Hu and Bentler (1999) recommended using a SRMR cutoff value close to .08 in conjunction with a RMSEA cutoff value close to .06. Following this two-index presentation strategy, the original five-factor model met the criterion of model fit (SRMR = .065, RMSEA = .064), whereas the three-factor and one-factor models were rejected (SRMR = .070, .076, respectively; RMSEA = .074, .099, respectively). Furthermore, there was no overlap between the RMSEA values and the 90% RMSEA confidence intervals (CIs) of the five-factor and three-factor model with the one-factor model (see Table 3). This supports a multidimensional approach for the measurement of masculine gender role stress.

TABLE 1
Intercorrelations Between MGRS Subscales

Subscale	*1*	*2*	*3*	*4*	*5*
1. Physical Inadequacy	—	.48*	.56*	.58*	.58*
2. Emotional Inexpressiveness		—	.43*	.60*	.39*
3. Subordination to Women			—	.57*	.39*
4. Intellectual Inferiority				—	.58*
5. Performance Failure					—

Note. N = 2,023. MGRS = Masculine Gender Role Stress.
*p < .01, two-tailed.

TABLE 2
Intercorrelations Between FGRS Subscales

Subscale	*1*	*2*	*3*	*4*	*5*
1. Unemotional Relationships	—	.60*	.62*	.57*	.65*
2. Physical Unattractiveness		—	.58*	.50*	.55*
3. Victimization			—	.60*	.53*
4. Behaving Assertively				—	.56*
5. Not Being Nurturant					—

Note. N = 2,023. FGRS = Feminine Gender Role Stress.
*p < .01, two-tailed.

Next, additional fit indexes were examined (see Table 3). With regard to the comparative fit index (CFI), no model met the criterion of model fit of CFI ≥ .95 (Hu & Bentler, 1999). However, the five-factor and three-factor models showed higher CFI values than the one-factor model. The relative fit indexes, expected value of the cross-validation index (ECVI), Akaike's (1987) information criterion (AIC), and population discrepancy function (PDF) indicated a consistent pattern of poorer fit for the one-factor model relative to the multidimensional factor models. In addition, the three-factor model revealed more favorable relative fit indexes than the original five-factor model. The ranges of the 90% CIs for ECVI and PDF showed no overlap for the three different factor models, indicating no interdependence between the three competing models.

Comparison of the fit indexes of the MGRS models for women were in agreement with those of men. There were virtually no cross-sex differences on the SRMR and RMSEA values of the multidimensional models. Cross-sex differences between model fit were more marked for the one-factor model. In addition, the CFI, the relative fit indexes, and the 90% CIs showed similar patterns of cross-sex differences on the factor models (see Table 3).

With regard to the FGRS scale, comparison of the fit indexes of the different models revealed practically similar results to those yielded in testing the different MGRS models. Based on Hu and Bentler's (1999) two-index presentation strategy for SRMR and RMSEA, the original five-factor model met the criterion of model fit (SRMR = .070, RMSEA = .063). The three-factor and one-factor models were rejected (SRMR = .084, .070, respectively; RMSEA = .081, .088, respectively).

In relation to the additional fit indexes (see Table 3), none of the models was supported by the CFI. It is worth noting that the one-factor model showed the poorest fit. Similar to the MGRS scale, comparisons of the relative fit indexes (ECVI, AIC, and PDF) indicated lack of support for the

TABLE 3
Fit Indexes by Scale, Sex, and Factor Model

Scale	*SRMR*	*RMSEA*	*90% CI*	*CFI*	*ECVI*	*90% CI*	*AIC*	*PDF*	*90% CI*
MGRS									
Men									
Five-factor model[a]	.065	.064	.062 to .067	.80	4.08	3.87 to 4.29	3,346	2.97	2.76 to 3.19
Three-factor model[b]	.070	.074	.069 to .078	.81	1.35	1.23 to 1.47	1,104	1.01	0.89 to 1.14
One-factor model	.076	.099	.097 to .101	.64	8.36	8.04 to 8.68	6,860	7.26	6.94 to 7.58
Women									
Five-factor model[a]	.064	.067	.065 to .069	.73	4.01	3.84 to 4.20	4,815	3.25	3.08 to 3.44
Three-factor model[b]	.070	.071	.067 to .074	.81	1.15	1.06 to 1.25	1,385	0.92	0.83 to 1.02
One-factor model	.069	.085	.083 to .087	.61	6.10	5.78 to 6.33	7,321	5.35	5.13 to 5.58
FGRS									
Men									
Five-factor model[c]	.062	.060	.057 to .063	.82	3.52	3.33 to 3.72	2,891	2.47	2.27 to 2.66
Three-factor model[b]	.079	.089	.082 to .093	.83	1.28	1.16 to 1.41	1,051	1.02	0.91 to 1.15
One-factor model	.061	.077	.075 to .079	.71	5.21	4.97 to 5.46	4,279	4.17	3.93 to 4.42
Women									
Five-factor model[c]	.070	.063	.061 to .065	.78	3.46	3.30 to 3.63	4,155	2.74	2.58 to 2.91
Three-factor model[b]	.084	.081	.076 to .085	.82	1.03	0.95 to 1.13	1,240	0.86	0.77 to 0.95
One-factor model	.070	.088	.086 to .089	.60	6.09	5.87 to 6.32	7,312	5.39	5.16 to 5.61

Note. Men: n = 822; women: n = 1,201. SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; CI = confidence interval; CFI = comparative fit index; ECVI = expected value of cross-validation index; AIC = Akaike's information criterion; PDF = population discrepancy function; MGRS = Masculine Gender Role Stress; FGRS = Feminine Gender Role Stress.
[a]Eisler and Skidmore (1987). [b]Tang and Lau (1996). [c]Gillespie and Eisler (1992).

one-factor model relative to the three-factor and five-factor model. The three-factor model also reflected more favorable relative fit indexes than the five-factor model. Furthermore, the 90% CIs for ECVI and PDF revealed no interdependence between the different factor models.

In accordance with the results on the MGRS scale, the SRMR and RMSEA values of the multidimensional models of the FGRS scale showed practically no cross-sex differences, whereas cross-sex differences on the one-factor model were more marked. Furthermore, the CFI, the relative fit indexes, and the 90% CIs reflected similar patterns of cross-sex differences on the factor models (see Table 3).

Reliability

The internal consistency coefficients for both gender role stress scales are presented in Table 4. In addition, Table 4 shows for each gender role stress scale the homogeneity index and the range of the item-remainder correlations. Cronbach's alpha for the total MGRS scale was high (.90). In relation to the MGRS subscales, Cronbach's alphas for Subordination to Women, Performance Failure, and Physical Inferiority were satisfactory (.80, .76, and .70, respectively), whereas Cronbach's alphas for Emotional Inexpressiveness and Intellectual Inferiority were lower (both .69) but acceptable for research purposes if the mean interitem correlation is also taken into account (Streiner, 2003). With regard to the total FGRS scale, Cronbach's alpha was high (.93), and the Cronbach's alphas for the FGRS subscales were satisfactory (ranging from .77 to .81). The mean interitem correlation and the range of the item-remainder correlations were for all gender role stress (sub)scales well within acceptable to satisfactory limits.

Further Analyses of Construct Validity

Table 5 shows means and standard deviations on the MGRS and FGRS scales and subscales for female and male students separately. In addition, Table 5 shows results of analyses of variance for sex differences. No sex differences were found on the total MGRS scale and the MGRS Emotional Inexpressiveness and Subordination to Women subscales. On the remaining MGRS subscales, however, sex differences were found. That is, compared to female students, male students scored significantly higher on the Physical Inadequacy subscale but significantly lower on the Intellectual Inferiority and Performance Failure subscales. Furthermore, on the FGRS total scale and subscales, sex differences were found in the predicted direction, that is, female students had higher FGRS scores than male students.

As anticipated, scores on the MGRS and FGRS scales were positively correlated to scores on the daily hassles scale (r = .40, $p < .001$; and r = .39, $p < .001$, respectively).

Furthermore, the MGRS scale showed no significant correlations with the masculinity–femininity scales except for the Masculinity scale of the PAQ (r = –.21). In addition, the FGRS scale was positively correlated to the femininity scales (r = .20 and r = .22, BSRI and PAQ, respectively) and negatively correlated to the masculinity scales (r = –.26 and r = –.28, BSRI and PAQ, respectively; all ps ≤ .001).

Hotelling's t tests for differences between dependent correlations revealed that the correlations between gender role stress and daily hassles were significantly stronger than the correlations between gender role stress and masculinity–femininity (t values ranging from 2.22 to 3.81, all ps < .05).

TABLE 4
Internal Consistency (Cronbach's α), Homogeneity (Mean Interitem Correlation), and Range of Item-Remainder Correlations of the MGRS and FGRS Scales and Subscales

Scale	*Cronbach's α*	*M Interitem Correlation*	*Range Item-Remainder Correlations*
MGRS	.90	.19	.24 to .54
Physical Inadequacy	.70	.20	.27 to .46
Emotional Inexpressiveness	.69	.24	.29 to .56
Subordination to Women	.80	.32	.36 to .63
Intellectual Inferiority	.69	.23	.25 to .49
Performance Failure	.76	.29	.34 to .54
FGRS	.93	.26	.38 to .60
Unemotional Relationships	.81	.30	.42 to .58
Physical Unattractiveness	.80	.34	.40 to .65
Victimization	.79	.38	.37 to .61
Behaving Assertively	.81	.38	.41 to .66
Not Being Nurturant	.77	.30	.38 to .53

Note. N = 2,023. MGRS = Masculine Gender Role Stress Scale; FGRS = Feminine Gender Role Stress Scale.

TABLE 5
Means and Standard Deviations of the Scores on MGRS and FGRS Scales and Subscales for Female and Male Students and Results for Analyses of Variance for Sex Differences

	Women[a]		*Men*[b]			
Scale	*M*	*SD*	*M*	*SD*	*F*	*Cohen's d*
MGRS	1.84	0.54	1.81	0.63	1.08	–0.05
Physical Inadequacy	1.81	0.66	2.00	0.77	37.57*	0.27
Emotional Inexpressiveness	1.64	0.77	1.62	0.78	0.19	–0.03
Subordination to Women	0.88	0.63	0.88	0.75	0.03	0.00
Intellectual Inferiority	1.73	0.71	1.57	0.81	21.54*	–0.21
Performance Failure	3.21	0.70	3.00	0.86	36.99*	–0.27
FGRS	2.92	0.62	2.19	0.70	619.66*	1.12
Unemotional Relationships	2.99	0.76	2.15	0.83	547.47*	1.06
Physical Unattractiveness	2.60	0.88	1.83	0.85	383.71*	0.89
Victimization	2.85	0.88	1.89	0.93	549.72*	1.07
Behaving Assertively	2.74	0.89	2.19	0.95	170.63*	0.60
Not Being Nurturant	3.36	0.68	2.80	0.79	286.72*	0.77

Note. MGRS = Masculine Gender Role Stress; FGRS = Feminine Gender Role Stress.
[a]n = 1,201. [b]n = 822.
*p < .001, one-tailed.

DISCUSSION

In general, the findings of this study validate the usefulness of the Dutch gender role stress scales. They verify the factorial validity and the cross-sex factorial invariance of the dimensions underlying the MGRS and FGRS scales as well as the internal consistency and homogeneity of both gender role stress scales. Although the MGRS scale was not sex specific, further evidence for construct validity showed that both gender role stress scales were correlated with a measure of daily hassles and were, to a lesser extent, correlated with measures of masculinity–femininity.

With regard to the factorial validity of the Dutch gender role stress scales, the original five-factor structures were supported. Overall, the CFAs showed a distinctively better fit for the five-factor models (Eisler & Skidmore, 1987; Gillespie & Eisler, 1992) and the three-factor models (Tang & Lau, 1996) relative to the one-factor models. This finding supports a multidimensional approach of the assessment of masculine and feminine gender role stress and is congruent with previous research. Based on the two-index presentation strategy (Hu & Bentler, 1999) and the content of the multidimensional models, the original five-factor structures of the gender role stress scales are preferred to the three-factor solutions. With respect to cross-sex factorial invariance, previous research reported identical factor structures in men and women on the MGRS scale (McCreary, Newcomb, & Sadava, 1998). Extending these findings, this study showed on both the original MGRS and FGRS scales practically similar fit indexes for Dutch male and female students. These similarities indicate that male and female students interpreted the factor structures in the same way.

Previous studies on the gender role stress scales have demonstrated that the scales are sex specific. That is, com-

pared to women, men reported more stress in situations that require masculine attributes but less stress than women in the situations that require feminine characteristics (Eisler & Skidmore, 1987; Gillespie & Eisler, 1992; Tang & Lau, 1995). In line with these results, this study showed that Dutch female students reported more FGRS than Dutch male students, both on the FGRS total and subscales. However, this study failed to find that the MGRS scale is sex specific. Dutch male students did not experience significantly greater levels of MGRS than Dutch female students, except for the Physical Inadequacy subscale. Besides, sex differences in the opposite direction to that expected were found on the Intellectual Inferiority and Performance Failure subscales. Overall, on the MGRS scale, men scored below the scale mean, whereas women scored above the scale mean on the FGRS scale. This difference might indicate that Dutch men reported a little masculine gender role stress, whereas Dutch women reported some real feminine gender role stress.

In addition, the averaged gender role stress scores of this Dutch sample were, separate for each sex, compared to those of the American and Chinese samples (Eisler & Skidmore, 1987; Gillespie & Eisler, 1992; Tang & Lau, 1995). Scores from Eisler and Skidmore (1987) were prorated because of different scaling (6-point vs. 7-point Likert scale). Dutch students reported less masculine gender role stress than American and Chinese students (t values ranging from 4.80 to 9.94, all ps $< .01$). Dutch students also reported less stress than American and Chinese students when faced with challenges to the feminine gender role (t values ranging from 2.61 to 10.77, all ps $< .01$).

An explanation for the finding that Dutch male students did not report more stress in situations that require masculine characteristics than Dutch female students might be a cultural difference with regard to masculinity–femininity. According to Hofstede (1991) masculinity–femininity is one of four independent dimensions that distinguishes national cultures. In a study conducted in 50 countries and three regions, Hofstede (1991) demonstrated that the Dutch live in a relative feminine society in which gender roles overlap, whereas Americans live in a relative masculine society in which gender roles are clearly distinct. Based on this distinction, men in the United States are supposed to be more assertive and focused on success and have been programmed with tougher values relative to men in The Netherlands. The finding, that men did not report higher levels of MGRS than women, might thus be explained by an overlap between the gender roles of men and women in The Netherlands.

The construct validity of the Dutch gender role stress scales was first supported by the previously mentioned CFAs. Further evidence of construct validity was found with regard to the correlations. First, this study revealed significant positive correlations between gender role stress and daily hassles, indicating that both instruments measure an aspect of stress (Gillespie & Eisler, 1992). Second, this study showed no significant correlations between the measures of MGRS and masculinity–femininity (BSRI and PAQ) except for a significant negative correlation between the MGRS scale and the Masculinity scale of the PAQ. In addition, the FGRS scale revealed significant positive correlations with the femininity scales and significant negative correlations with the masculinity scales of the BSRI and PAQ. Previous research demonstrated no association between MGRS and Masculinity (Eisler & Skidmore, 1987) and no correlation between FGRS and Femininity (Gillespie & Eisler, 1992). This lack of association supported the distinction between the concepts of gender role stress and masculinity–femininity. The significant correlations between the measures of gender role stress and masculinity–femininity found in this study all revealed small to medium effect sizes, whereas the correlations between the measures of gender role stress and daily hassles were stronger with medium to large effect sizes (Cohen, 1992). Taken together, the pattern of correlations found in this study supported, although not conclusive, the construct validity of the gender role stress scales.

A limitation of this study is that only students were included. The extent to which this sample differs from the general Dutch population limits the generalizability of the results. Future research needs to explore the validity and reliability of the gender role stress scales among a more heterogeneous group of participants. Furthermore, the use of different methods of data collection could be argued. To avoid any differences due to mode of administration, the format of the paper and computer version of the questionnaire were held as similar as possible (Webster & Compeau, 1996). Nevertheless, there could be some discussion on the online data collection. Participants could have completed the online questionnaire several times or could have asked another person to complete the questionnaire on their behalf. However, there is a growing body of evidence that Internet-based data collection can produce similar results to the more traditional methods of data collection (e.g., Epstein, Klinkenberg, Wiley, & McKinley, 2001). In addition, analysis on the MGRS and FGRS averaged scale scores adjusting for sex, age, and discipline did not show an overall significant effect of method of data collection. Moreover, additional analyses for each data collection method separately resulted in similar conclusions about the factorial validity and reliability of both gender role stress scales.

In summary, the findings of this study validate the usefulness of the Dutch version of the gender role stress scales in The Netherlands. These instruments enable researchers to further investigate the associations between gender role stress and vulnerability to health problems.

REFERENCES

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Arrindell, W. A., Kolk, A. M., Pickersgill, M. J., & Hageman, W. J. (1993). Biological sex, sex role orientation, masculine sex role stress, dissimula-

tion and self-reported fears. *Advances in Behaviour Research and Therapy, 15,* 103–146.

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

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross cultural research* (Vol. 8, pp. 137–164). Beverly Hills, CA: Sage.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

de Jong, G. M., Timmerman, I. G. H., & Emmelkamp, P. M. G. (1996). The survey of recent life experiences: A psychometric evaluation. *Journal of Behavioral Medicine, 19,* 529–542.

Eisler, R. M. (1995). The relationship between masculine gender role stress and men's health risk: The validation of a construct. In R. F. Levant & W. S. Pollack (Eds.), *A new psychology of men* (pp. 207–225). New York: Basic.

Eisler, R. M., & Blalock, J. A. (1991). Masculine gender role stress: Implications for the assessment of men. *Clinical Psychology Review, 11,* 45–60.

Eisler, R. M., & Skidmore, J. R. (1987). Masculine gender role stress: Scale development and component factors in the appraisal of stressful situations. *Behavior Modification, 11,* 123–136.

Eisler, R. M., Skidmore, J. R., & Ward, C. H. (1988). Masculine gender-role stress: Predictor of anger, anxiety, and health-risk behaviors. *Journal of Personality Assessment, 52,* 133–141.

Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior, 17,* 339–346.

Gillespie, B. L., & Eisler, R. M. (1992). Development of the feminine gender role stress scale: A cognitive-behavioral measure of stress, appraisal, and coping for women. *Behavior Modification, 16,* 426–438.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles, 7,* 1097–1108.

Hofstede, G. (1991). *Cultures and organizations: Software of the mind.* London: McGraw-Hill.

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

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

Huisman, M. (1999). Simple and effective methods to treat missing item responses. *Kwantitatieve Methoden, 62,* 57–78.

Jöreskog, K., & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language [Computer software]. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kohn, P. M., & Macdonald, J. E. (1992). The Survey of Recent Life Experiences: A decontaminated hassles scale for adults. *Journal of Behavioral Medicine, 15,* 221–236.

Kolk, A. M., & de Ruiter, C. (1993). *Nederlandse vertaling Bem Sex Role Inventory* [Dutch translation of the Bem Sex Role Inventory]. Unpublished manuscript.

Martz, D. M., Handley, K. B., & Eisler, R. M. (1995). The relationship between feminine gender role stress, body image, and eating disorders. *Psychology of Women Quarterly, 19,* 493–508.

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

McCreary, D. R., Newcomb, M. D., & Sadava, S. W. (1998). Dimensions of the male gender role: A confirmatory analysis in men and women. *Sex Roles, 39,* 81–95.

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

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

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha. *Journal of Personality Assessment, 80,* 99–103.

Tang, C. S., & Lau, B. H. (1995). The assessment of gender role stress for Chinese. *Sex Roles, 33,* 587–595.

Tang, C. S., & Lau, B. H. (1996). The Chinese gender role stress scales: Factor structure and predictive validity. *Behavior Modification, 20,* 321–337.

van Well, S., & Kolk, A. M. (2002). *Nederlandse vertaling Personal Attributes Questionnaire* [Dutch translation of the Personal Attributes Questionnaire]. Unpublished manuscript.

Webster, J., & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. *Behavior Research Methods, Instruments and Computers, 28,* 567–576.

Sonja van Well
Department of Clinical Psychology
University of Amsterdam
Roetersstraat 15
1018 WB Amsterdam
The Netherlands
Email: S.M.vanWell@uva.nl

Received April 15, 2004
Revised September 23, 2004