TPM®

PRIMARY CARE PHYSICIANS AND GENDER MEDICINE: VALIDATION OF THE ITALIAN VERSION OF NIJMEGEN GENDER AWARENESS IN MEDICINE SCALE (N-GAMS)

SILVIA GATTINO
GIORGIA MOLINENGO
NORMA DE PICCOLI
UNIVERSITY OF TORINO, ITALY

The psychometric properties of the Nijmegen Gender Awareness in Medicine Scale (N-GAMS) are evaluated by confirmatory factor analysis (CFA) and its invariance is tested using multigroup procedures. Four hundred and thirty-one Italian physicians (44.5% women; M_{age} = 50.1) completed the Italian version of N-GAMS: 311 were primary care physicians and 120 were medical students specializing in primary care. The CFA revealed that the N-GAMS consisted of three factors: gender sensitivity (GS), gender role ideology toward patients (GRI-P), and gender role ideology toward doctors (GRI-D). Removal of two items (1 and 13) from the GS factor improved the fit, so that the final version of the Italian scale consists of 30 items. Multigroup analysis showed that N-GAMS was metric invariant between women and men, whereas it achieved scalar invariance between in-service physicians and postgraduate medical students.

Keywords: Gender awareness; Gender medicine; Confirmatory factor analysis; Invariance; Practicing physicians.

Correspondence concerning this article should be addressed to Giorgia Molinengo, Department of Psychology, University of Torino, Via Verdi 10, 10124 Torino (TO), Italy. Email: giorgia.molinengo@unito.it

Gender bias in health care underlies health disparities (Hamberg, 2008; Humphries et al., 2017; Morais et al., 2020). There is a broad empirical basis in epidemiological data for evaluating health from a gender perspective: the same diseases that affect men and women differ in frequency, symptoms, prognosis, and response to treatment. Gender also affects medical communication (Dielissen et al., 2012): how patients present symptoms (Gattino et al., 2019; Danielsson et al., 2009; Hall & Roter, 2002) and how physicians interpret complaints and signs of illness (Andersson et al., 2012; Chiaramonte & Friend, 2006; Hamberg et al., 2002).

The world's leading health organizations have set the goal of improving gender equality in health through various initiatives, statements, and policies. The World Health Organization (WHO, 2015) has defined gender as part of its 2014-2019 programming. An aim of sex gender equity in research (Heidari et al., 2016) is to include information about the sex and gender of medical trial participants in all study phases. In addition, gender is a key issue in the formulation of the 2030 agenda and its sustainable development goals (SDGs), with the inclusion of a stand-alone development global goal (SDG5) to achieve gender equality and women's rights.

Despite these advances, male bias in medicine still holds sway. Knowledge about health and disease predominantly refers to men and men's health, with the result that men and women are systematically and erroneously viewed as either the same or different (Risberg et al., 2009; Ruiz-Cantero et al., 2007), with repercussions for the health of women and men alike.

Gender Awareness in Health Care

In general terms, taking gender into account in health care means aiming for appropriateness, that is, more effective and efficient interventions in prevention, diagnosis, treatment, and rehabilitation (Andersson et al., 2012). To achieve this goal, health care workers need to be aware of gender-related aspects of health and how to tailor medical attention and promote greater gender equality in health care. Gender awareness in health care means that physicians know how to recognize gender and incorporate it into their daily practice as a determinant of health and disease. Previous studies (Verdonk et al., 2007, 2008) investigating gender awareness among medical students found that the female students generally took a more patient-centered approach, which correlated with less adherence to conventional stereotypes about men and women and greater gender sensitivity (Verdonk et al., 2007). The professional role models to which medical students aspired also varied: the ideal physician for male students was primarily associated with hierarchical authority, decision making, rationality, competitiveness, and objectivity, whereas female students associated the ideal physician with caring, concern for others, and understanding. In a more recent study involving students in medicine and health care, Siller and colleagues (2018) underlined the importance of gender medicine in the course curriculum. Gender sensitivity, also among laboratory staff, was positively associated with gender teaching. Furthermore, male students were noted to benefit more than female students from taking courses in gender medicine.

A review of the scientific literature published between 1967 and 2001 reported that, compared to their male counterparts, female physicians provide more information, are more attentive, and show more interest in patients' emotions, lifestyle, and family ties (Roter et al., 2002). A recent study by Gattino and colleagues (2020), investigating awareness of the gender dimension among family physicians and postgraduate medical students, found that the women were more sensitive to gender issues than their male colleagues and that the professional role contributed to gender awareness: male and female postgraduate medical students were more sensitive to addressing gender differences in their practice.

Miller and colleagues first defined the concept of gender awareness in health care in 1999. The various instruments to operationalize this construct differ in what they measure and what they describe. These differences reflect the debate surrounding the concept of gender awareness (Miller et al., 1999; Khoury & Weisman, 2002; Verdonk et al., 2009). Two validated instruments to assess gender awareness on a theoretical basis in health care workers are the Gender Awareness Inventory-Veterans Administration (GAI-VA; Salgado et al., 2002) and the Nijmegen Gender Awareness in Medicine Scale (N-GAMS; Verdonk et al., 2008). The GAI-VA scale is based on the model of gender awareness proposed by Miller and colleagues (1999) and was developed and validated for a population of U.S. military veterans, that is, a social group composed mainly of men. The GAI-VA assesses three dimensions of gender awareness among health care workers: gender sensitivity, gender ideology, and knowledge. Gender sensitivity indicates the extent to which health care workers are aware of and understand the needs of women patients. Gender ideology indicates the attitude of medical staff toward these veterans, while knowledge refers to the accuracy of information about patients and their needs. The scale has demonstrated good psychometric properties, but its focus on this patient category precludes its extension to other health care contexts.

Differently, the N-GAMS assesses gender awareness among medical students toward their patients in general and extends it to male and female physicians, thus overcoming one of the limitations of the GAI-VA. The N-GAMS also assesses three dimensions: gender sensitivity (GS), gender role ideology toward patients (GRI-P), and gender role ideology toward doctors (GRI-D). GS refers to the extent to which medical students are able to perceive gender differences and the impact of gender on medical practice; GRI-P refers to stereotypes about how male and female patients communicate and cope with health and illness; and GRI-

D refers to gender stereotypes of physicians. The three subscales have shown good reliability and the scale has shown good criterion validity as a whole, indicating that the N-GAMS is a useful tool to measure gender awareness among health care workers.

The N-GAMS is currently one of the main instruments developed to assess the level of gender awareness of health professionals and has been used in several studies and in different contexts. Andersson and colleagues (2012) used it to assess and compare the gender awareness of Dutch and Swedish medical students; Eisenberg and colleagues (2013) used it to examine the impact of a women's health intervention program on the gender awareness of a group of Australian medical students in their final year of study. This scale has also been used to examine gender awareness in medical and allied health students at the University of Tyrol and the University of Insbruck (Siller et al., 2018), to assess gender awareness in a sample of French-Swiss medical students (Rrustemi et al., 2020), and used in Italy to assess gender awareness of general practitioners and physicians in training (Gattino et al., 2020) and to compare attitudes toward gender and gender stereotypes among male and female medical students (Barbadoro et al., 2022; Bert et al., 2022). These studies demonstrated the applicability of the N-GAMS in different contexts to assess how gender stereotyping may affect medical training and practice and the effectiveness of educational programmes aimed at increasing gender awareness in health care.

This scale is not only widely used, but has also been validated in other European and non-European contexts. In particular, Morais and colleagues (2020) adapted and validated it on a sample of Portuguese medical students. Their results confirmed the construct validity, reliability, sensitivity, and 3-factor structure of the N-GAMS. Results also showed that more empathic students reported higher GS and lower agreement with GRI-P and GRI-D; higher hostile and benevolent sexism were associated with higher endorsement of GRI-P and GRI-D, and higher hostile sexism with lower GS. However, as the authors themselves point out, one of the limitations of this study is that it was conducted on a sample of medical students with little or no practical experience. Aliri and colleagues (2022) adapted the N-GAMS to the Spanish population of nurses in service and in training and analyzed its psychometric properties. This study confirmed the 3-factor structure of this instrument and its validity and reliability in assessing the attitudinal component of gender awareness not only among physicians and medical students but also among nurses and nursing students. Finally, an Arabic contextualization of N-GAMS has also recently been proposed (Shamasneh et al., 2023). The researchers presented this scale to a sample of Palestinian primary care physicians and nurses. Shamasneh and colleagues partially confirmed the 3-factor structure of the N-GAMS and showed a difference between men and women in the subscales GRI-P and GRI-D. In both subscales, women expressed fewer stereotypical beliefs than men.

Given the relevance of the gender dimension in health care, developing valid instruments that can reliably measure physician gender awareness is a critical step in supporting the claim that increased physician gender awareness helps minimize gender bias in health care. In addition, such instruments can be advantageously employed to evaluate the effectiveness of intervention programs for increasing gender awareness among health care workers. Drawing on these assumptions and the literature described, the main aim of the current study was to work toward this goal by adapting and validating the Nijmegen Gender Awareness Scale (Verdonk et al., 2008) to the Italian context. This study builds on a previous study that examined the extent to which general pratictioners (GPs) and GPs in training are aware of gender medicine and the extent to which gender stereotypes are perpetuated by GPs and their colleagues in training (Gattino et al., 2020). Our aim is different from that of the aforementioned study. The aim of this study is to validate the N-GAMS in the Italian context. To this end, we tested the psychometric properties of the N-GAMS and extended the study to the criterion-related validity of the measure by comparing experienced and nonexperienced GPs. This comparison allows us to overcome some limitations of previous studies and to extend the scientific knowledge on measurement of gender awareness in medicine.

Aims and Hypotheses

With the present study we evaluated the psychometric properties of the N-GAMS in a sample of Italian general practitioners and postgraduate medical students specializing in general medicine. To the best of our knowledge, there are no validated instruments to assess gender awareness among Italian health care workers. Our study had three aims. The first was to test the factorial structure of the scale by confirmatory factor analysis (CFA). We wanted to understand whether the answers to the 32 original items were governed by a single latent dimension (unidimensional model), by two dimensions — one referring to gender sensitivity and the other to gender-role ideology (bidimensional model) — or by the 3-factor structure (tridimensional model) originally proposed by Verdonk et al. (2008) — one referring to gender sensitivity (GS), and the other two measuring gender role ideology toward patients (GRI-P) and gender role ideology toward doctors (GRI-D). Our hypothesis was that the tridimensional model would show a better fit to the data than either the unidimensional or the bidimensional model (Hypothesis 1).

The second aim was to determine, using a multigroup analysis, the invariance of the confirmatory factor model between men and women and between practicing physicians and medical students in postgraduate specialization. Our hypothesis was that the model is invariant (Hypothesis 2) in relation to these variables and allows for an appropriate and meaningful comparison of N-GAMS values related to gender and professional role (Meredith, 1993; Meredith & Horn, 2001). By determining invariance, we assumed that mean differences indicated true mean differences and not measurement bias, which is critical in the context of measure validation (Coulacoglou & Saklofske, 2017). Therefore, our third aim was to obtain descriptive statistics for each subscale of the N-GAMS in order to determine the level of gender awareness of sample members and any differences by gender or professional role. Following Verdonk et al's (2008) discussion of gender differences in gender awareness, we expected that female physicians would have higher levels of gender sensitivity and lower levels of gender role ideologies than their male counterparts. Regarding differences between GPs and GPs in training, we might expect that the still predominant biomedical model in medical education combined with less experience might decrease gender awareness in GPs in training. On the other hand, increased attention to gender issues and the inclusion of ad hoc curricular courses might increase gender awareness in medical students. In light of these conflicting expectations, our aim was to explore the presence of any differences or similarities between general practitioners and postgraduate medical students.

Method

Procedure

The study procedure adhered to the ethical standards set by the Italian Psychological Association and the principles of the 1964 Helsinki Declaration. The research project was approved (n. 60196) by the Research Ethic Committee, University of Torino (Italy), and the study was not preregistered.

Participants were recruited through a convenience sampling methods and data were collected by the researchers and by research assistants trained by the researchers. The study entailed administering a structured, self-report, pencil-paper questionnaire that took approximately 20 minutes to complete. Participation was voluntary and was not compensated. Respondent anonymity was guaranteed at all stages of data collection and analysis. The general practitioners completed the questionnaire at the beginning of their monthly

team meetings with one of the researchers present and the postgraduate medical students completed it at the beginning or end of one of their lectures. Questionnaires were collected immediately after completion.

Participants

The study sample was 431 Italian physicians (44.5% women; age range 27-72 years, *M* = 50.14, *SD* = 13.49): 311 (41.2% women) were primary care physicians (age range 30-72 years, *M* = 57.63, *SD* = 6.67) and 120 (53% women) were postgraduate medical students specializing in primary care (age range 27-54 years, *M* = 30.79, *SD* = 4.14). The two groups did not differ by gender, $\chi^2(1) = 4.75$, $p = .20$, which is consistent with the fact that in Italy women account for 44% of the total number of medical school graduates, and for as much as 64% in the 35-to 59-year-old age group (CERGAS-Bocconi, 2019).

Measure

Nijmegen Gender Awareness in Medicine Scale (N-GAMS). To adapt and validate the N-GAMS to the Italian sample, we followed international guidelines on the adaptation and cross-cultural validation of instruments for measuring psychological constructs (Beaton et al., 2000). The N-GAMS consists of three subscales: gender sensitivity (GS; 14 items), gender role ideology toward patients (GRI-P; 11 items), and gender role ideology toward doctors (GRI-D; 7 items). The gender sensitivity subscale assesses the degree to which physicians are sensitive to dealing with gender issues (e.g., "Physician's knowledge of gender differences in illness and health increase quality of care"). This measure demonstrated good internal consistency (α = .85, ω = .86). The gender role ideology toward patients and toward physicians evaluates the care with which physicians avoid stereotypical generalizations toward patients and physicians, respectively. For example: "Male patients understand the approach of physicians better than female patients"; "Male physicians put more emphasis on the technical aspects of medicine than female physicians." Both measures demonstrated good internal consistency (GRI-P, α = .88, ω = .88; GRI-D, α = .78, ω = .77). Response options for all three subscales range from 1 (*strongly disagree*) to 5 (*strongly agree*).[1]

Data Analysis

Statistical analysis was performed using IBM SPSS Statistics Version 27 and its MATRIX language, and the Mplus 8 analysis program. Confirmatory factor analysis (CFA) was conducted using multiple linear regression (MLR) estimation (Muthén & Muthén, 2004) because the variables showed a marked violation of normal multivariate distribution (Mardia's kurtosis coefficient 1293.56, $p < .000$), and the sample was large enough.

Model evaluation and comparison were conducted using standard goodness-of-fit indices: comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root-mean-square residual (SRMR), and root-mean-square error of approximation (RMSEA) with associated confidence interval (CI). Based on the work of Hu and Bentler (1999), CFI and TLI values of .95 or higher are considered as indicating good model-data fit, and values of .90 and < .95 are taken as acceptable fit; SRMR values equal to .08 or lower, and RMSEA values equal to .06 and lower would indicate a good fit between the hypothesized model and the data. Modification indeces (MIs) were also used to identify the parameters responsible for the lack of fit in the measurement models.

A multigroup CFA was conducted to assess measurement invariance, and four measurement invariance steps were specified according to Putnick and Bornstein (2016): (1) configural, equivalence of model form; (2) metric (weak factorial), equivalence of factor loadings; (3) scalar (strong factorial), equivalence of item intercepts or thresholds; (4) residual (strict or invariant uniqueness), equivalence of items' residuals or unique variances. The goodness of fit of each model was compared to that of the previous one throught the Satorra and Bentler chi-square difference test ($\Delta\chi^2$) (Satorra & Bentler, 2001; see also Bryant & Satorra, 2012). According to Chen (2007), the following changes in goodness-of-fit indices were considered indicative of a lack of invariance: criterion of a –.01 change in CFI, paired with changes in RMSEA of .01 and SRMR of .03 (for metric invariance) or .02 (for scalar invariance). One-way multiple analysis of variance (one-way MANOVA) was used to analyze mean differences on the three N-GAMS scores for both gender and professional role.

Results

To probe the underlying N-GAMS structure, 1-, 2-, and 3-factor solutions were determined (Table 1). The fit for the unidimensional and the bidimensional model were inadequate. CFI and TLI indices were too low and both RMSEA and SRMR were far from acceptable. The 3-dimensional model was empirically more plausible and reproduced the theoretical articulation of the three N-GAMS subscales (Table 1), with one factor related to gender sensitivity, a second to gender role ideology toward patients, and a third to gender role ideology toward doctors. MIs were all low and suggest no further changes to the estimated model, hovewer two items (1, 13) on gender sensitivity were not significant, with a factor loading equal to .21 and .25, respectively, and were deleted from the model. The solution showed an increase in fit with acceptable CFI (.94), RMSEA (.05), and SRMR (.06) indices.

Table 1
Confirmatory factor analysis

Model	χ^2 MLR (*df*)	CFI	TLI	RMSEA [90% CI]	SRMR
Unidimensional	2212.03 (464)	.54	.51	.09 [.09, .10]	.12
Bidimensional	1192.66 (463)	.81	.79	.10 [.10, .11]	.10
Tridimensional	989.86 (461)	.92	.90	.05 [.05, .06]	.06
Tridimensional_modified*	851.22 (402)	.94	.92	.05 [.05, .06]	.06

Note. MLR = multiple linear regression; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual. *Model after deletion of Items 1 and 13.

The item saturations in the respective factors were similarly high (Figure 1) and all were statistically significant ($p < .05$). The standardized loadings ranged from .43 to .75 for GS, from .45 to .78 for GRI-P, and from .41 to .70 for GRI-D. The interfactor correlation was also high and statistically significant ($p < .05$). The higher correlation was between GRI-P and GRI-D ($r = .66$) and GS showed a negative relation both with GRI-P ($r = –.28$) and GRI-D ($r = –19$).

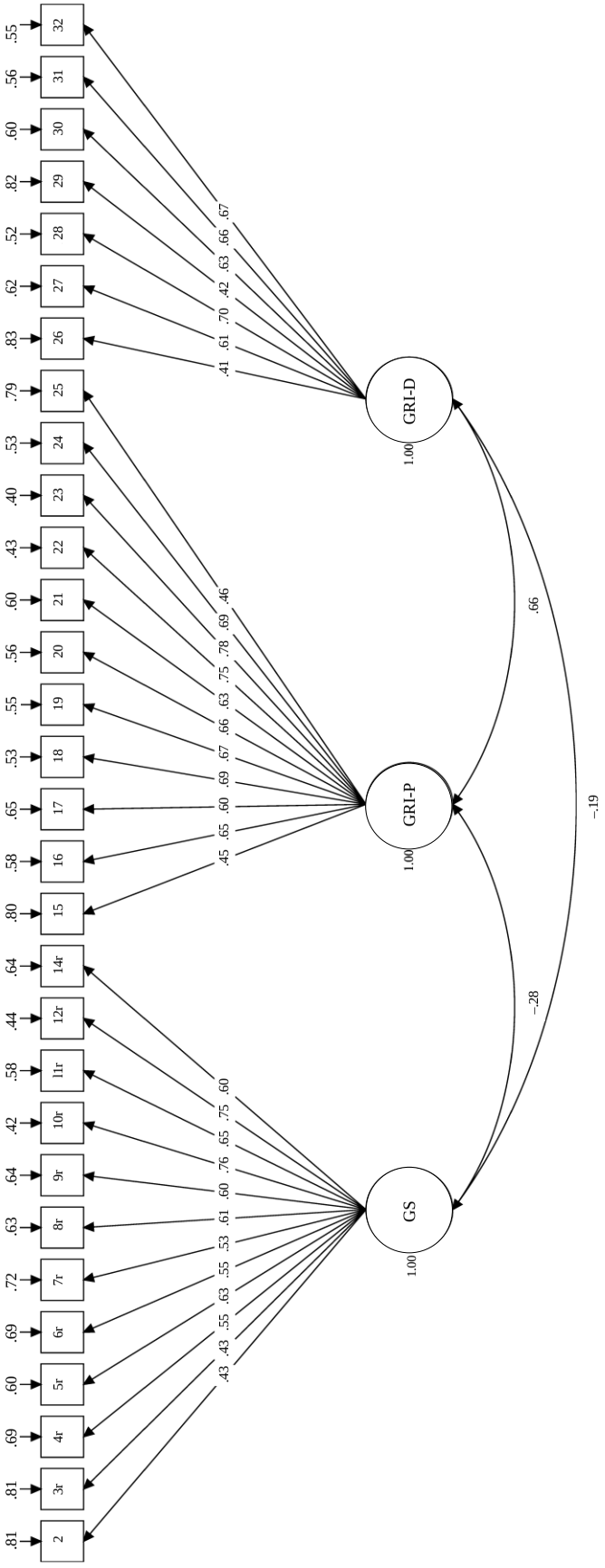

FIGURE 1
Three correlated factor model (standardized solution)

Note. GS = gender sensitivity; GRI-P = gender ideology toward patients; GRI-D = gender ideology toward doctors.

The 3-dimensional factor model was subjected to structural invariance analysis for gender and professional roles (Table 2). As mentioned in Data Analysis, the process proceeded by multiple steps. The configural model was estimated without imposing intergroup equality restrictions on the parameters. The configural model was used both to check that the number of factors in each group was equal and to test the various invariance hypotheses. Once we were sure that the manifest variables were indicators of the same factors in the samples, we could delve deeper into whether the structure can be generalized.

Table 2
Measurement invariance analysis for gender and for professional role

Invariance model	$\chi^2(df)$	RMSEA	CFI	SRMR	$\Delta\chi^2$	Δdf	Δ RMSEA	Δ CFI	Δ SRMR
Gender (men, *n* =239; women, *n* = 192)									
Configural	1420.67(804)	.06	.94	.07	–	–	–	–	–
Metric	1429.85(831)	.06	.94	.07	18.66	27	–.00	.00	.00
Scalar	1532.76(858)	.06	.91	.08	112.85*	27	.00	–.03	.01
Residual	1455.52(864)	.11	.83	.15	–	–	–	–	–
Professional role (primary care physicians, *n* = 311; postgraduate medical students, *n* =120)									
Configural	1476.53(804)	.06	.94	.07	–	–	–	–	–
Metric	1511.81(831)	.06	.93	.08	37.69	27	.00	–.01	.01
Scalar	1586.57(858)	.06	.92	.08	77.21*	27	.01	–.01	.00
Residual	1728.14(864)	.07	.79	.12	–	–	–	–	–

Note. *df* = degrees of freedom; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; SRMR = standardized root-mean-square residual; $\Delta\chi^2$ = Satorra-Bentler χ^2 difference test.
*$p < .05$.

With respect to gender, the configural model was compared with the metric one, which imposed the constraint of equality of male and female factor saturations in addition to equality of factor number. The Satorra-Bentler chi-square difference test showed no significant difference between the two models and changes in CFI, RMSEA, and SRMR were adeguate. Comparison between the scalar model, which imposed equality of the intercepts, and the less restricted metric model returned several unsatisfactory results: the change in CFI was above –.01, paired with a chi-square difference test ($\Delta\chi^2$) significant; finally, the invariance of the residual model did not fit the data well for CFI, RMSEA, and SRMR.

For professional role, the fit of the configural, metric, and scalar models was good, whereas the most restricted one (the residual model) showed improper values for the fit indices. Therefore, we can state that the N-GAMS has a more than acceptable level of invariance for both gender and professional role: the number of dimensions and the relationship between them and the indicators were, indeed, the same. The differences were in the item intercepts and unique variances between men and women, whereas for professional role only unique variances varied between GPs and GPs in training.

Table 3 shows the mean values of the three N-GAMS dimensions in relation to gender and professional role and MANOVA was used to test for differences across groups. Previous results have shown that scalar invariance is established for professional role, thus allowing for a meaningful comparison of the mean scores of these groups, while the achievement of metric invariance with respect to gender suggested that

caution should be exercized in meaningfully comparing scores between males and females. With this in mind there was a statistically significant difference in N-GAMS subscales based on gender — $F(3, 380) = 3.59$, $p = .01$; Wilks $\Lambda = .97$, partial $\eta^2 = .03$ — but men and women diverged significantly only with respect to GS: female physicians were more likely than their men colleagues to perceive gender differences and the influence of gender on medical practice; no statistically significant differences emerged with respect to GRI-P and GRI-D.

Table 3
Sample descriptives using one-way MANOVA for equality of means

	Gender					Professional role				
	Women		Men		$F(1, 382)$	Primary care physicians		Postgraduate medical students		$F(1, 391)$
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	
GS	3.91	0.73	3.70	0.87	8.48**	3.65	0.86	4.04	0.64	12.32**
GRI-P	2.23	0.77	2.37	0.82	*ns*	2.36	0.82	2.19	0.75	2.07*
GRI-D	2.04	0.64	2.04	0.76	*ns*	2.05	0.73	1.98	0.71	*ns*

Note. GS = gender sensitivity; GRI-P = gender role ideology toward patients; GRI-D = gender role ideology toward doctors.
*$p < .05$; **$p < .01$.

With regard to professional role, the MANOVA results — $F(3, 389) = 6.45$, $p < .01$; Wilks $\Lambda = .95$, partial $\eta^2 = .05$ — revealed that there are statistical differences for GS and GRI-P: primary care physicians showed a lower ability to perceive gender differences and were more inclined to stereotype the way male and female patients communicate and deal with health and illness. Finally, differences in GRI-D were not significant between primary care physicians and post graduate medical students specializing in primary care.

Discussion and Conclusions

The aim of the present study was to analyze the psychometric properties of the Italian version of the Nijmegen Gender Awareness in Medicine Scale (N-GAMS; Verdonk et al., 2008). Consistent with our first hypothesis, confirmatory factor analysis showed that the 3-factor model fit the data better than the other models. As in the original study (Verdonk et al., 2008), our findings showed that the N-GAMS consists of three interrelated factors: gender sensitivity, gender role ideology toward patients, and gender role ideology toward doctors. Items 1 and 13 of gender sensitivity showed low and not significant factor loading and were removed from the model. The final version of the Italian scale consists of 30 items and has more than adequate fit and validity. The 3-factor structure was structurally stable over all professional roles and gender subgroups, thus supporting our second hypothesis.

Multigroup analysis yielded results well above the values required to consider the scale invariant between in-service physicians and postgraduate medical students, whereas N-GAMS was metric invariant for gender only. Therefore, the number of dimensions and all factorial saturations were the same for both genders and professional roles. The differences were in item intercepts and unique variances between men

and women, whereas for professional role only unique variances varied between general practitioners and general practitioners in training.

Overall, our results show adequate psychometric properties for construct validity of the N-GAMS, but suggest that its tripartite structure allows adequate and meaningful comparisons of scores for professional roles, whereas such comparisons with respect to gender must be made with caution. However, as Putnik and Bornstein (2016) suggested, noninvariance can be instructive and lead researchers to important insights into how different groups perceive the same construct, and failure to demonstrate invariance does not necessarily preclude further analysis of group differences.

The results of mean differences on the three subscales of N-GAMS showed that GS scores were higher for female physicians and postgraduate medical students, suggesting that these two subgroups are more likely to perceive gender differences and the influence of gender on medical practice. Regarding the GRI-P subscale, primary care physicians were more inclined to stereotype the way male and female patients communicate and cope with health and illness than their colleagues in training, whereas there were no differences between men and women. Finally, on the GRI-D subscale, there were no differences between men and women or between general practitioners and postgraduate medical students specializing in primary care.

This study has several limitations. Our findings cannot be generalized to all Italian physicians because of a convenience sampling method and the resulting lack of representativeness. Although the sample was sufficiently large for statistical analysis, further studies are needed to better understand whether geographic variance exists between different parts of Italy. Also, the sample consisted of general practitioners either in service or in training. It would be interesting to compare studies involving physicians in other specialties.

These limitations aside, the present study has several strengths. It confirms the underlying 3-factor structure of the N-GAMS in a sociocultural context different from the original one. In this version, the scale appears valid and reliable for assessing gender awareness among health care workers. It is also easy to administer, complete, and score. Our study provides an opportunity to extend current knowledge about the psychometric qualities of the N-GAMS to assess gender awareness in physicians in training and shows that the Italian version of the N-GAMS appears to be a robust psychological instrument for use in research and clinical interventions.

Note

1. The Italian version of the scale is available upon request from the first author.

References

Aliri, J., Prego-Jimenez, S., Goñi-Balentziaga, O., Pereda-Pereda, E., Perez-Tejada, J., & Labaka Etxeberria, A. (2022). Gender awareness is also nurses' business: Measuring sensitivity and role ideology towards patients. *Journal of Nursing Management, 30*(8), 4409-4418.https://doi.org/10.1111/jonm.13866

Andersson, J., Verdonk, P., Johansson, E. E., Lagro-Janssen, T., & Hamberg, K. (2012). Comparing gender awareness in Dutch and Swedish first-year medical students — Results from a questionnaire. *BMC Medical Education, 12*(3), 1-10. https://doi.org/10.1186/1472-6920-12-3

Barbadoro, P., Ponzio, E., Dolcini, J., Ricciardi, A., Di Tondo, E., & D'Errico, M. M. (2022). Gender awareness among Italian medical students. *Italian Journal of Gender-Specific Medicine, 8*(3), 135-142.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186-3191. https://doi.org/10.1097/00007632-200012150-00014

Bert, F., Boietti, E., Rousset, S., Pompili, E., Franzini Tibaldeo, E., Gea, M., Scaioli, G., & Siliquini, R. (2022). Gender sensitivity and stereotypes in medical university students: An Italian cross-sectional study. *PLoS One*, *17*(1), Article e0262324. https://doi.org/10.1371/journal.pone.0262324

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(3), 372-398. https://doi.org/10.1080/10705511.2012.687671

CERGAS-Bocconi. (2019). *Rapporto OASI* 2019 [*OASI Report*]. Retrieved from https://cergas.unibocconi.eu/sites/default/files/files/Cap2OASI_2019.pdf

Chiaramonte, G. R., & Friend, R. (2006). Medical students' and residents' gender bias in the diagnosis, treatment, and interpretation of coronary heart disease symptoms. *Health Psychology*, *25*(3), 255-266. https://doi.org/10.1037/0278-6133.25.3.255

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Coulacoglou, C., & Saklofske, D. H. (2017). *Psychometrics and psychological assessment: Principles and applications*. Academic Press.

Danielsson, U., Bengs, C., Lehti, A., Hammarström, A., & Johansson, E. (2009). Struck by lightning or slowly suffocating — Gendered trajectories into depression. *BMC Family Practice*, *10*(56), 1-11. https://doi.org/10.1186/1471-2296-10-56

Dielissen, P., Verdonk, P., Bottema, B., Kramer, A., & Lagro-Janssen, T. (2012). Expert consensus on gender criteria for assessment in medical communication edcuation. *Patient Education and Counseling*, *88*(2), 189-195. https://doi.org/10.1016/j.pec.2012.01.013

Eisenberg, K., Dahlstrom, T., Carnovale, A. J., Neeman, T., & Ellwood, D. (2013). Gender awareness in medical curriculum: Surveying final year students undertaking a women's health rotation. *Medical Teacher*, *35*(11), 970-971. https://doi.org/10.3109/01421 59x.2013.78681 8

Gattino, S., De Piccoli, N., Grosso, M., Miozzo, S., & Tanturri, G. (2019). Gender awareness in medicine, sexism and empathy. A comparison between professionals and general practitioners in training. *Psicologia della Salute*, *3*, 7-27. https://doi.org/10.3280/PDS2019-003002

Gattino, S., De Piccoli, N., Grosso, M., Miozzo, S., Tanturri, G., & Rollero, C. (2020). Awareness of gender medicine among family doctors. A field investigation, *Journal of Prevention and Intervention in the Community*, *48*(2), 147-160. https://doi.org/10.1080/10852352.2019.1624354

Hall, J. A., & Roter, D. L. (2002). Do patients talk differently to male and female physicians? A meta-analytic review. *Patient Education and Counseling*, *48*(3), 217-224. https://doi.org/10.1016/S0738-3991(02)00174-X

Hamberg, K. (2008). Gender bias in medicine. *Women's Health*, *4*(3), 237-243. https://doi.org/10.2217/17455057.4.3.237

Hamberg, K., Risberg, G., Johansson, E. E., & Westman, G. (2002). Gender bias in physicians' management of neck pain: A study of the answers in a Swedish national examination, *Journal of Women's Health & Gender-Based Medicine*, *11*(7), 653-666. https://doi.org/10.1089/152460902760360595

Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review*, *1*(2), 1-9. https://doi.org/10.1186/s41073-016-0007-6

Hu, L., & 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. https://doi.org/10.1080/10705519909540118

Humphries, K. H., Izadnegadar, M., Sedlak, T., Saw, J., Johnston, N., Schenck-Gustafsson, K., Shah, R. U., Regitz-Zagrosek, V., Grewal, J., Vaccarino, V., Wei, J., & Bairey Merz, C. N. (2017). Sex differences in cardiovascular disease — Impact on care and outcomes. *Frontiers in Neuroendocrinology*, *46*, 46-70. https://doi.org/10.1016/j.yfrne.2017.04.001

Khoury, A. J., & Weisman, C. S. (2002). Thinking about women's health: The case for gender sensitivity. *Women's Health Issues*, *12*(2), 61-65. https://doi.org/10.1016/S1049-3867(01)00146-3

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525-543. https://doi.org/10.1007/BF02294825

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203-240). American Psychological Association. https://doi.org/10.1037/10409-007

Miller, P., Vogt, L., King, J., & King, D. W. (1999). Gender ideology, sensitivity and knowledge: A model of gender awareness in VA health care. *Veterans Health System Journal*, *4*, 27-30.

Morais, R., Bernardes, S. F., & Verdonk, P. (2020). Gender awareness in medicine: Adaptation and validation of the Nijmegen Gender Awareness in Medicine Scale to the Portuguese population (N-GAMS). *Advances in Health Sciences Education*, *25*, 457-477. https://doi.org/10.1007/s10459-019-09936-y

Muthén, L. K., & Muthén, B. (2004). *MPLUS user's guide*. Muthén & Muthén.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71-90.

Risberg, G., Johansson, E. E., & Hamberg, K. (2009). A theoretical model for analysing gender bias in medicine. *International Journal for Equity in Health*, *8*(28), 1-8. https://doi.org/10.1186/1475-9276-8-28
Roter, D. L., Hall, J. A., & Aoki, Y. (2002). Physician gender effects in medical communication. A meta-analytic review, *Journal of American Medical Association*, *288*(6), 756-764. https://doi.org/10.1001/jama.288.6.756
Ruiz-Cantero, M., Vives-Cases, C., Artazcoz, L., Delgado, A., Calvente, M. D. M., Miqueo, C., Montero, I., Ortiz, R., Ronda, E, Ruiz, I., & Valls, C. (2007). A framework to analyse gender bias in epidemiological research. *Journal of Epidemiology & Community Health*, *61* (Suppl. 2), 46-53. https://doi.org/10.1136/jech.2007.062034
Rrustemi, I., Locatelli, I., Schwarz, J., Lagro-Janssen, T., Fauvel, A., & Clair, C. (2020). Gender awareness among medical students in a Swiss university. *BMC Medical Education*, *20*, 1-8. https://doi.org/10.1186/s12909-020-02037-0
Salgado, D. M., Vogt, D. S., King, L. A., & King, D. W. (2002). Gender Awareness Inventory-VA: A measure of ideology, sensitivity, and knowledge related to women veterans' health care. *Sex Roles*, *46*(7), 247-262. https://doi.org/10.1023/A:1020171416038
Satorra, A., & Bentler, P. M. (2001). A scaled difference χ^2 test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514. https://doi.org/10.1007/BF02296192
Siller, H., Komlenac, N., Fink, H., Perkhofer, S., & Hochleitner, M. (2018). Promoting gender in medical and allied health professions education: Influence on students' gender awareness. *Health Care for Women International*, *2*, 1-17. https://doi.org/10.1080/07399332.2017.1395881
Shamasneh, B., Nemer, M., & Abu-Rmeileh, N. M. (2023). Gender awareness in healthcare: Contextualization of an Arabic version of the Nijmegen Gender Awareness in Medicine Scale (N-GAMS). *Healthcare*, *11*(4), Article 629. http://dx.doi.org/10.3390/healthcare11040629
Verdonk, P., Benschop, Y., de Haes, H., & Lagro-Janssen, T. (2008). Medical students' gender awareness Construction of the Nijmegen Gender Awareness in Medicine Scale (N-GAMS). *Sex Roles*, *58*, 222-234. https://doi.org/10.1007/s11199-007-9326-x
Verdonk, P., Benschop, Y., de Haes, H., Mans, L., & Lagro-Janssen, T. (2009). Should you turn this into a complete gender matter? Gender mainstreaming in medical education. *Gender and Education*, *21*(6), 703-719. https://doi.org/10.1080/09540250902785905
Verdonk, P., Harting, A. J., & Lagro-Janssen, T. L. M. (2007). Does equal education gender equal attitudes? Gender differences in medical students' attitudes toward the ideal physician. *Teaching and Learning Medicine, 19*, 9-13. https://doi.org/10.1080/10401330709336617
World Health Organization. (2015). *Integrating equity, gender, human rights and social determinants into the work of WHO*. World Health Organization.