Research paper

Psychometric properties of the EURO-D scale of depressive symptomatology: Evidence from SHARE wave 8

José M. Tomás [a], Zaira Torres [a,*], Amparo Oliver [a], Sara Enrique [b], Irene Fernández [a]

[a] *Department of Methodology for the Behavioral Sciences, University of Valencia, Valencia, Spain*
[b] *Department of Developmental Psychology and Education, University of Valencia, Valencia, Spain*

A R T I C L E   I N F O

A B S T R A C T

Background: The EURO-D is a short scale to measure symptoms of depression, very used in large population surveys. Although there are numerous validation studies, its psychometric properties remain unclear. The two-factor structure (Affective Suffering and Lack of Motivation) is replicated in several studies but with different item compositions, and none reported reliability indices for both factors. For that reason, the aim of this study is to examine the factorial validity of the scale, the reliability of the dimensions, the gender differential item functioning (DIF), and the nomological validity.
Methods: 46,317 participants aged 50 and over (M = 71.33), from which 57.4 % were females, in Wave 8 of the Survey of Health, Aging and Retirement in Europe (SHARE) were included. Instruments: EURO-D, R-UCLA, Self-perceived health index, and indicator of taking drugs for anxiety or depression. Factor Analyses, DIF, Reliability Index and Spearman correlations were estimated.
Results: Factor analysis identified a bifactor structure: a general factor of Depression and two specific factors of Affective Suffering and Lack of Motivation, which reliabilities were 0.83, 0.83 and 0.79 respectively. No relevant DIF item by gender was found, but higher scores were found in women in all factors. Both factors had positive relations with loneliness, taking drugs and self-perceived health.
Limitations: this study has a cross-sectional design; future research may consider the longitudinal stability of the scale.
Conclusions: EURO-D shows adequate psychometric properties when a general factor of depression and two specific factors are considered. Women have higher scores on all dimensions.

1. Introduction

Depression is a debilitating and globally common disease; it differs from normal mood swings by the extent of its severity, symptoms, and duration (World Health Organization, 2012). Indeed, depression is the mental health problem with the highest worldwide prevalence (Wu et al., 2020). According to the Global Health Data Exchange (Institute for Health Metrics and Evaluation, October 2021), it is estimated that 280 million Europeans have suffered depression. This would mean that depression is affecting 3.8 % of the population, an estimate that increases with age, with 5 % of the adults and 5.7 % of the old adults (>60 years old) suffering from it. The high prevalence of depression symptoms is concerning and suggests the need for routine investigation of the construct by professionals (Akosile et al., 2018). Furthermore, the

diagnosis in the population of old adults is complex due to its confusion with frequent life situations such as grief, moving home, loss of physical and mental capacities, among other problems (Alexopoulos, 2005).

There are several well-established scales for the measurement of depressive symptoms that are showed in Table 1. For population surveys, self-administered short scales are more suitable. From these, the Center for Epidemiologic Studies—Depression Scale (CES—D) of 20-item (Markush and Favero, 1973; Radloff, 1977), which is systematically used in its briefer 8-item form in both the English Longitudinal Study of Aging (ELSA) and the Health and Retirement Survey (HRS) in the USA. Finally, the one we focus on in this study is the EURO-D scale (Prince et al., 1999), a self-report questionnaire originally developed to compare the symptoms of depression in old age in 11 European countries (Prince et al., 1999). The EURO-D scale was developed merging

items from these scales: GMS, SHORT-CARE, ZSDS and the CPRS (Courtin et al., 2015). The EURO-D has been administered in a large, longitudinal, household survey of representative samples of people aged 50 years and over from European countries: the Survey of Health, Aging and Retirement in Europe (SHARE).

Regarding the factor structure of the EURO-D, previous research suggests a solution of two factors. In a first study, Prince et al. (1999) found a structure with two factors in people aged 65 years and over (N = 21,724) in fourteen European countries. They found consistently across countries that the first factor, Affective Suffering, was characterized by items related to depression, tearfulness, and suicidality, accompanied by lower loadings from items related to pessimism, appetite, and fatigue. The second factor, Motivation, had content related to loss of interest, poor concentration, and lack of enjoyment. This structure was replicated using the first wave of the SHARE data (N = 22,777) by Castro-Costa et al. (2008) with Confirmatory Factor Analysis (CFA) in people aged 50 years or older from ten European countries. Based on Prince et al. (1999) results, Castro-Costa et al. (2008) also found the same two factors, but with modifications. The same two-factor structure found in the study of Guerra et al. (2015) employing a CFA in a sample of 17,852 old people, aged 65 years or more, from nine non-European countries (Cuba, Dominican Republic, Puerto Rico, Peru, Mexico, Venezuela, China, India, and Nigeria). In another study, Portellano-Ortiz et al. (2018) with a sample of 62,182 participants aged 65 of over from the fifth wave of SHARE data in 15 European countries, found a similar structure to the previous studies, but discarding the suicide item because it did not load onto any factor. Recently, Maskileyson et al. (2021) proved the same two-factor structure in 41,862 participants, aged 50 or older, from the sixth wave of SHARE data in 14 European countries plus Israel using Exploratory Factor Analysis (EFA) and CFA, with similar factor loadings to those in Portellano-Ortiz et al. (2018), but excluding several items, among them suicidality that cross-loaded. In sum, the factor structure of the EURO-D scale remains unclear, with many doubts about the role suicidality play.

Regarding reliability of the scale, although all the aforementioned studies supported the two-factor structure, none of them reported the reliabilities of the two dimensions of the EURO-D, only some studies informed about it for the full scale. Prince et al. (1999) in their first EURO-D study reported a Cronbach's alpha value ranging between 0.58 and 0.80 across countries. In the study of Castro-Costa et al. (2008), the EURO-D scale had alphas ranging from 0.62 to 0.78, and in the study of

Guerra et al. (2015) between 0.64 and 0.87. For the EURO-D validation in Spain, Larraga et al. (2006) found an alpha value of 0.75 and in the Thailand's sample, Jirapramukpitak et al. (2009) a value of 0.72. In other study, Courtin et al. (2015) using data from the second SHARE wave had a Cronbach's alpha value of 0.72. Portellano-Ortiz et al. (2018) had an alpha of 0.71. There were no estimates of reliability in the study by Maskileyson et al. (2021).

About the diagnosis validity, many studies agree that a higher cut-off point of 5/6 on the scale would improve its specificity (Guerra et al., 2015; Jirapramukpitak et al., 2009; Prajwal et al., 2021) compared to the cut-off point of 3/4 originally established (Prince et al., 1999). EURO-D has showed evidence of good relations with criteria and with its nomological net of variables. For example, the scale has been positively and significantly correlated with loneliness, poor self-perceived physical health, being female and low cognition (Portellano-Ortiz and Conde-Sala, 2018), disability (Guerra et al., 2015; Prajwal et al., 2021) and inversely with happiness (Guerra et al., 2015).

Finally, gender seems to play a role in depression. One of the most consistent findings in epidemiology of mental health is the gender gap in depression (van de Velde et al., 2010), a gender gap that has also been found recently in European old adults (Schmitz and Lazarević, 2020). Both EURO-D factors were related to gender, but gender differences were larger in Affective Suffering, with women scoring higher (Brailean et al., 2015; Castro-Costa et al., 2007; Portellano-Ortiz and Conde-Sala, 2018; Prince et al., 1999), while the Motivation factor is more related to age and poor cognition (Portellano-Ortiz and Conde-Sala, 2018; Prince et al., 1999). To guarantee meaningful gender comparisons of depressive symptomatology, non-differential item functioning of the scale must be previously established. However, there are no gender invariance published studies of the EURO-D.

The aim of this study is to examine the psychometric properties of the EURO-D scale with data coming from all countries measured in wave 8 of SHARE. Specifically, we shed light on a) factorial validity of the scale, b) reliability of the dimension(s), c) gender differential item functioning, and d) nomological validity of the scale, using loneliness, self-perceived health and taking drugs for anxiety or depression as criterions.

Table 1
Characteristics of the instruments measuring depressive symptoms.

Instrument	Authors	Construct	Target population	Mode of administration	Items
Hamilton Rating Scale for Depression (HRSD)	Hamilton (1960)	Depressive symptoms	Diagnosed population	Semi-structured standardized interview scale	17
Geriatric Mental State Schedule (GMS)	Copeland et al. (1976)	Mental state	Diagnosed old adults	Semi-structured standardized interview scale	541
SHORT-CARE scale created from the Comprehensive Assessment and Referral Evaluation (CARE)	Gurland et al. (1984)	Depression, Dementia, and Disability	Old adults	Semi-structured standardized interview scale	143
Comprehensive Psychopathological Rating Scale (CPRS)	Asberg and Schalling (1979)	Broad psychiatric symptoms	Diagnosed population	Semi-structured standardized interview scale	65
Shorter version of the Comprehensive Psychopathological Rating Scale (CPRS-S-A)	Svanborg and Åsberg (1994)	Depression, Anxiety and Obsessional symptoms	Diagnosed population	Self-report	19
Beck Depression Inventory (BDI)	Beck et al. (1961)	Depressive symptoms	Non-diagnosed and diagnosed populations	Self-report	21
Zung Self-Rating Depression Scale (ZSDS)	Zung (1965)	Depressive symptoms	Diagnosed population	Self-report	20
Geriatric Depression Scale (GDS)	Sheikh and Yesavage (1986)	Depressive symptoms	Old adults	Self-report	30
Shorter version of the Geriatric Depression Scale (GDS—S)	Sheikh and Yesavage (1986)	Depressive symptoms	Old adults	Self-report	15
Center for Epidemiologic Studies—Depression Scale (CES—D)	Markush and Favero (1973); Radloff (1977)	Depressive symptoms	Old adults	Self-report	8
EURO-D scale	Prince et al. (1999)	Depressive symptoms	Old adults	Self-report	12

2. Method

2.1. Sample and procedure

Data used in this study comes from wave 8 of the Survey of Health, Aging and Retirement in Europe (SHARE; Börsch-Supan, 2022; Börsch-Supan et al., 2013). The SHARE project is aimed at populations aged 50 or more across several European countries and Israel. It follows a probability-based sampling strategy and presents a longitudinal design, with wave 8 representing the most recently gathered data in the project. The sampling protocol in SHARE involves four stages. First, a sample frame and procedure are specified by each participating country. Second, information about the sample is gathered and processed in each country in order to produce a preliminary sample file. Third, international coordinators from SHARE headquarters revises and approves the files. Finally, sample files are loaded onto the software system and combined with address information. Further details about sampling can be consulted at Bethmann et al. (2019). The SHARE study is continuously ongoing ethics review. Wave 8 data recollection was approved by the Ethics Council of the Max Plank Society (an overview can be found here).

The present study involved 46,317 participants aged between 50 and 104 years old ($M = 71.33$, $SD = 4.16$), from which 26,591 (57.4 %) were females and 19,726 (42.6 %) were males. Most of them were either married (57.5 %), widowed (15.6 %) or divorced (7.2 %), while the rest (19.7 %) presented other marital status. Sample size in each country is shown in Fig. 1.

2.2. Instruments

Depression was measured with the EURO-D scale (Prince et al., 1999), which screens for the presence of twelve different symptoms of depression with twelve items, one per symptom: depressed mood, pessimism, suicidality, guilt, sleep, (lack of) interest, irritability, (loss of) appetite, fatigue, (lack of) concentration, (lack of) enjoyment and tearfulness. Response scale was dichotomous, with 0 indicating absence and 1 indicating presence.

Loneliness was assessed with the Three-Item Loneliness Scale (Hughes et al., 2004), a short version of the R-UCLA Loneliness Scale developed by Russel et al. (1980). The shortened version used in this study contained three items tapping the frequency of feelings related to

lack of companionship, exclusion and isolation. Responses were coded in a three-point Likert scale (1 = hardly ever or never, 2 = some of the time, 3 = often). Loneliness scores were computed as the sum of the individual responses to each of the three items. Composite Reliability Index (CRI) estimate was 0.88 for the general factor of loneliness.

Self-perceived health evaluations were gauged using a single indicator of general health. This item asked respondents to rate their overall health in a five-point Likert scale, ranging from 1 (poor) to 5 (excellent). This indicator comes from the SF-36 questionnaire developed by Ware and Gandek (1998).

A single binary indicator divided the sample among those that take drugs for either anxiety or depression and those who do not.

2.3. Statistical analyses

A series of competitive Confirmatory Factor Analyses (CFA) models were tested in order to assess the factor structure of the EURO-D scale (Prince et al., 1999). First, a unidimensional structure was tested (Model 1), as a baseline parsimonious model. Model 2 hypothesized two factors of depressive symptomatology: Affective Suffering and Lack of Motivation. In this model Affective Suffering comprised the items about depressive mood, suicidality, guilt, sleep problems, irritability, fatigue, and tearfulness; while the second factor, (lack of) Motivation, contained the items about pessimism, lack of interest, loss of appetite, lack of concentration and lack of enjoyment. After testing this model, one additional model (Model 3) was tested in order to examine the relationship of the suicidality item with both factors. Therefore, Model 3 loaded the suicidality item onto both factors. Finally, and given the relatively high correlation between the two factors, a bifactor model was also tested. The bifactor model hypothesizes a general factor of Depression, onto which all items load, and two orthogonal (uncorrelated) specific factors of Affective Suffering and (lack of) Motivation.

Model fit of the CFAs was examined using several statistics and fit indexes: the chi-square statistic (χ^2), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Squared Root Mean Residual (SRMR). CFI values equal or higher than 0.90 and RMSEA and SRMR values equal or lower than 0.08 are deemed adequate (Hu and Bentler, 1999). Values of at least 0.95 for CFI and no >0.05 for RMSEA and SRMR are considered as indicating excellent fit (Hu and Bentler, 1999). To check relative improvements in fit among different models, given that models are not nested, we will use CFI
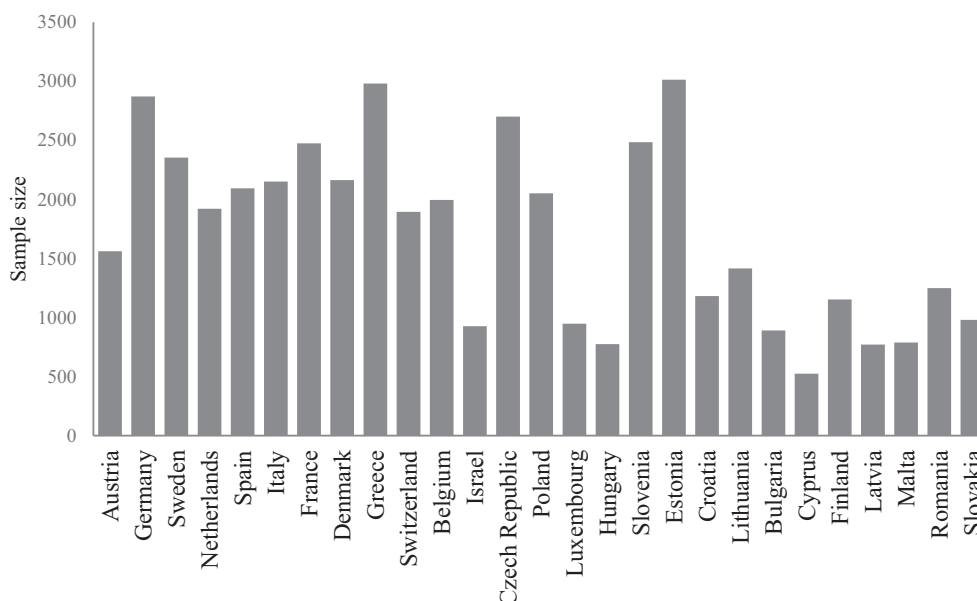

Fig. 1. Sample size in each country.

differences (ΔCFI). CFI differences larger than 0.01 indicate significant model deterioration (Cheung and Rensvold, 2002). All confirmatory models were estimated using Weighted Least Squares Mean and Variance corrected (WLSMV) as the method of estimation, given the binary nature of the items and their consequent deviation from multivariate normality (Finney and DiStefano, 2006). CFA models were estimated using MPlus 8.7 (Muthén and Muthén, 1998-2017). Complete cases were used for the analyses.

Once the factor structure of the scale was established, Differential Item Functioning (DIF) for men and women was analysed. Both uniform (different difficulties) DIF and non-uniform (different discrimination) DIF were studied with logistic regressions (Narayanan and Swaminathan, 1996). The presence of DIF was assessed statistically with a Likelihood Ratio Test (LRT), but given the extremely large sample size and considering the test too powerful, we also used Nagelkerke's R-square as a measure of effect size. There are two accepted guidelines to consider whether the effect sizes are negligible, the guidelines of Zumbo and Thomas (1997) (< 0.13 is negligible) and the ones by Jodoin and Gierl (2001) (< 0.035 is negligible). DIF analyses were performed in R (R Core Team, 2021) using the difR package (Magis et al., 2010).

Finally, Composite Reliability Index (CRI; Raykov, 2004) was computed to examine the reliability of the scale. To test convergent and discriminant validity of the EURO-D, Spearman correlations were computed to examine the relationship between depressive symptoms and three criteria: loneliness, self-perceived health and taking drugs for anxiety or depression.

3. Results

3.1. Factor structure

Fit indexes of the different CFAs are displayed in Table 2. The best fitting model in terms of all fit indices was Model 4, the bifactor model. Therefore, this is the model retained. Factor structure and loadings are shown in Fig. 2. All factor loadings were statistically significant, but given the extremely large sample size, this is very likely and tells not much. More important are the sizes of the standardized loadings. In the general factor of depression all, but two of them, are 0.4 or larger. The only two slightly lower than this cut-off are the ones associated to pessimism and lack of enjoyment. On the other hand, looking at the largest loadings in the specific factors give us an idea of which items retain more specific variance of affective suffering and lack of motivation. Regarding Affective Suffering: depressive mood, guilt irritability and tearfulness retain relatively large standardize loadings. In the same vein, pessimism, lack of interest and lack of enjoyment in the specific factor of Lack of Motivation also retain a large amount of specific variance.

3.2. Reliability and descriptive statistics

The CRI estimate for the general (depression) factor was 0.83, whereas the CRIs for the specific factors of Affective Suffering and Lack

Table 2
Fit indices of the tested models.

	Model 1 One factor	Model 2 Two factors	Model 3 Two-factors cross-loading	Model 4 Bifactor
χ^2	8505.28	4848.27	4383.56	1094.44
df	54	53	52	42
p	<0.001	<0.001	<0.001	<0.001
CFI	0.902	0.945	0.950	0.988
ΔCFI	–	0.045	0.005	0.086
RMSEA	0.059	0.045	0.043	0.023
90 % CI	0.085–0.060	0.044–0.046	0.042–0.044	0.022–0.025
SRMR	0.072	0.056	0.051	0.029

of Motivation were 0.49 and 0.46, respectively. Nevertheless, when the CRIs for these two factors are calculated with the results of Model 2, the estimates were 0.83 for Affective Suffering and 0.79 for Lack of Motivation. Table 3 shows mean and standard deviations for all the items. Given that items are binary, the mean is the percentage of the sample showing this symptom.

3.3. Differential item functioning by gender

DIF was assessed with two logistic regression analyses, one for detecting uniform DIF and another for detecting non-uniform DIF. The results of these statistics for all items and in both uniform and non-uniform DIF are presented in Table 3. The LRT statistics detected 8 items with uniform DIF and other 8 items with statistically significant non-uniform DIF. However, all DIF effect sizes were extremely low, and obviously negliglibe by both Zumbo and Thomas' criteria and by Jodoin and Gierl criteria. Therefore, there no evidence for relevant gender DIF.

DIF by gender was also assessed in each country separately. That is, uniform and non-uniform DIF was tested in each of the 27 countries. Some LRTs were statistically significant. Nevertheless, when R-squares were considered, no item showed uniform or non-uniform DIF in any country when using Zumbo and Thomas' criteria. However, there were a number of countries that exhibited moderate DIF in item 12 (tearfulness) by gender according to Jodoin and Gierl's criteria: Italy, Greece, Israel, Czech Republic, Hungary, Estonia, Croatia, Lithuania, Bulgaria, Cyprus, Latvia, Malta, Romania and Slovakia. Anyway, only Latvia had a R-square large than 0.1 (R-square = 0.17). Additionally, there was moderate DIF in item 2 (R-square = 0.04) for Finland. All in all, there is no relevant DIF in the scale across countries.

3.4. Gender differences in EURO-D factors

Regarding the overall depression factor a *t*-test show statistically significant differences with a low effect size ($t_{(44606)} = 36.06$; $p < .001$; Cohen's d = 0.35, 95 % CI (0.33, 0.36)). Concerning the relation between gender and the two EURO-D dimensions, independent *t*-tests were also calculated. There were differences between men and women in both, Affective Suffering: $t_{(44,077.415)} = 45.28$; $p < .001$; Cohen's d = 0.42, 95 % CI (0.64, 0.70), and Lack of Motivation: $t_{(42,691.706)} = 9.26$; $p < .001$; Cohen's d = 0.08, 95 % CI (0.07, 0.11). In both factors, women showed higher scores than men. Specifically, in Affective Suffering, women's mean was 2.04 (SD = 1.716) and men's mean was 1.36 (SD = 1.451), whereas in Lack of Motivation, women's mean was 0.73 (SD = 1.075) and men's mean was 0.64 (SD = 0.990). Finally, chi-square tests with Cramer's V as a measure of effect size were calculated to test for gender differences in the twelve items. Only items 2 and 11 showed statistically no significant gender differences ($p > .05$), but the effect sizes were of very low magnitude. The only Vs that were larger than 0.1 (the lower limit for low effect) were those of item 1 (depressive mood, V = 0.16), item 5 (sleeping problems, V = 0.15), and item 12 (tearfulness, V = 0.25).

3.5. Nomological validity

Spearman correlations of Affective Suffering and the selected criterions, loneliness, taking drugs for anxiety or depression and self-perceived health, were $r = 0.33$ ($p < .001$), $r = 0.23$ ($p < .001$), and $r = -0.38$ ($p < .001$), respectively. On its part, Lack of Motivation correlated $r = 0.29$ ($p < .001$) with loneliness, $r = 0.17$ ($p < .001$) with taking drugs for anxiety and depression, and $r = -0.33$ ($p < .001$) with self-perceived health. Additionally, we calculated the correlations of the criteria with the overall scale and the results were: $r = 0.44$ ($p < .001$) with loneliness, $r = 0.24$ ($p < .001$) with drugs for anxiety and depression, and $r = -0.43$ ($p < .001$) for self-perceived health.
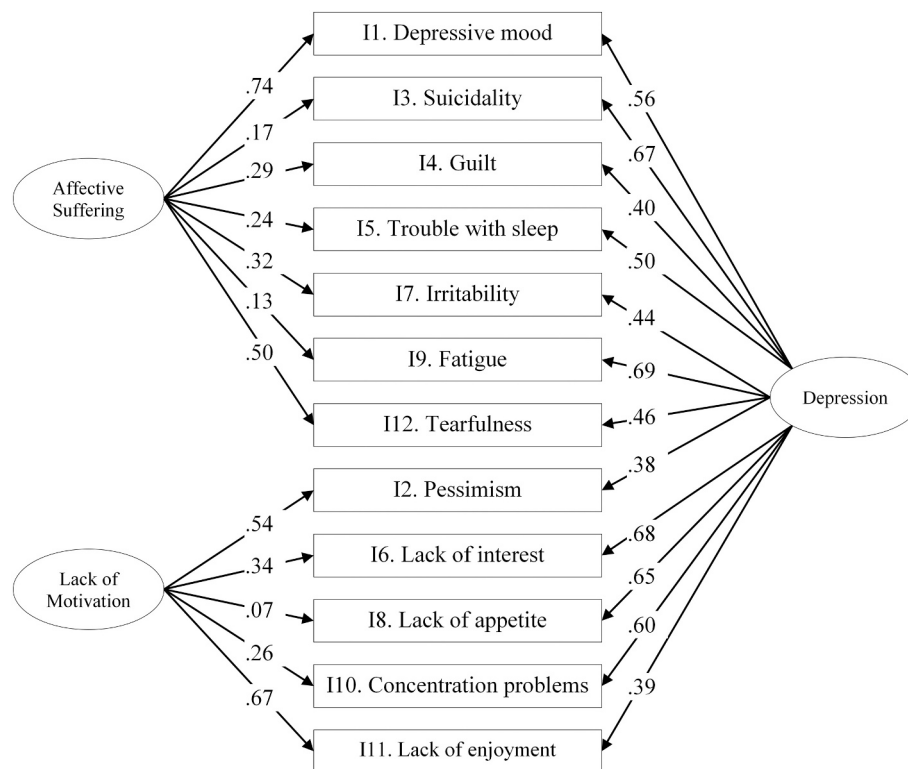
Fig. 2. Factor loadings of the retained model for EURO-D (Model 4).

Table 3

Mean (% of participants with the symptom), standard deviations, and uniform and non-uniform DIF statistics and Nagelkerke's R-squares with their *p*-values.

Item	Mean	SD	Uniform DIF			Non-uniform DIF		
			Statistic	*p*	R^2	Statistic	*p*	R^2
Depression	0.40	0.48	177.24	<0.001*	<0.001	0.602	0.43	<0.001
Pessimism	0.18	0.38	159.29	<0.001*	<0.001	57.73	<0.001*	<0.001
Suicidality	0.06	0.23	0.09	0.76	<0.001	0.00	0.95	<0.001
Guilt	0.07	0.26	1.37	0.24	<0.001	16.99	<0.001*	<0.001
Sleep	0.37	0.48	138.62	<0.001*	<0.001	17.52	<0.001*	<0.001
Interest	0.11	0.30	39.85	<0.001*	<0.001	3.05	0.08	<0.001
Irritability	0.26	0.43	202.86	<0.001*	<0.001	75.23	<0.001*	<0.001
Appetite	0.09	0.29	0.05	0.82	<0.001	0.66	0.41	<0.001
Fatigue	0.35	0.47	0.66	0.41	<0.001	17.37	<0.001*	<0.001
Concentration	0.17	0.37	125.18	<0.001*	<0.001	26.46	<0.001*	<0.001
Enjoyment	0.14	0.34	165.98	<0.001*	<0.001	37.24	<0.001*	<0.001
Tearfulness	0.24	0.42	1220.40	<0.001*	<0.001	76.01	<0.001*	<0.001

4. Discussion

Finding agreement in mental health measurement scales is very important, especially if they measure topics as relevant as depression and are used in large surveys, as is the case of EURO-D. Even though the structure of the EURO-D was tested in previous studies there is a lack of agreement, and the structure remains unclear. For that reason, the first objective of the study was analysing the factorial validity of the scale testing three models based on previous literature and an additional bi-factor model. The bifactor model clearly adjusted much better than any other structure. This complex, bifactor structure with both a strong general factor of depression and relevant specific variance in the factors of Affective Suffering and Lack of Motivation was never tested before. The fact that there is both a strong factor of Depression as well as specific variance in many items may partially explained why there were changes in the structure (with item either measuring their factor poorly, cross-loading or changing its pretended factor) in previous validation studies (Castro-Costa et al., 2008); Guerra et al., 2015; Maskileyson et al., 2021; Portellano-Ortiz et al., 2018; and Prince et al., 1999).

In our study all EURO-D items are well included in the general factor, while many items also retain specific and relevant variance, and this is important since the EURO-D was developed to capture the essence of the instruments GMS, SHORT-CARE, ZSDS and CPRS (Courtin et al., 2015; Prince et al., 1999), and excluding any of its items may damage the validity of the scale and its usefulness as a screening test.

Furthermore, our results clarify the contribution of each item to each factor, even in problematic items that have been excluded in previous studies. For example, the item about 'suicidality', excluded in the analysis of Portellano-Ortiz et al. (2018) and Maskileyson et al. (2021), has a good fit as a general item of Depression and has not very much specific variance. This may explain why in this study did not work well. Same happens with the 'appetite' item excluded in the studies of Castro-Costa et al. (2008) and Guerra et al. (2015). It loads highly on the global Depressive factor and very little on its specific factor (Lack of Motivation).

The second aim of the study was to analyze the reliability of the two

dimensions, Affective Suffering and Lack of Motivation, and also the general dimension. In all cases we found adequate internal consistency. Not only are our findings consistent with other studies that reported adequate EURO-D reliability (Courtin et al., 2015; Jirapramukpitak et al., 2009; Larraga et al., 2006; Portellano-Ortiz et al., 2018), but also, they are relevant because provided information of the reliability of the two dimensions separately.

Regarding the third aim of the study, to our knowledge, this is the first examination of gender-related DIF in EURO-D constructs. Detection of DIF is decisive because it can influence the instrument psychometric properties as well as mean score comparisons (Millsap, 2012). Although our results found statistically significant uniform/non-uniform DIF in 12 items, this finding is no surprise given the extremely large sample size employed in the study. Therefore, effect sizes based in Nagelkerke's R-squares were also calculated. There are two accepted guidelines to consider whether the effect sizes are negligible, the guidelines of Zumbo and Thomas (1997) and the ones by Jodoin and Gierl (2001). In both cases, all effect sizes, both for uniform and non-uniform DIF, were considered negligible with R-squares very close to zero. When DIF by gender was examined within each country, a moderate DIF only by the most restrictive criteria was found in several countries and mostly regarding item 12 (tearfulness). However, the magnitudes of DIF we found are negligible. Thence, we concluded there was no relevant gender-based DIF in the EURO-D scale with this wave 8 of the SHARE data. In other words, our findings suggest that, at the item level, the EURO-D constructs appear to equivalently function for women and men aged 50 or more across several European countries and Israel. Hence, EURO-D allows for meaningful and interpretable gender mean scores comparisons of depressive symptomatology.

Our results suggest that, actually, women tend to show higher Affective Suffering and Lack of Motivation compared to men, and these differences are not dependent on the EURO-D properties at the item level. Concretely, based on the effect sizes of these differences, Affective Suffering with a close to moderate effect size (Cohen's d higher than 0.4), was a more meaningful difference than the gender difference in Lack of Motivation, with a Cohen's d very low (Cohen, 1992). These findings are consistent with the previous literature, in which larger gender differences were found in Affective Suffering than in Lack of Motivation factor, with women scoring higher (Brailean et al., 2015; Castro-Costa et al., 2007; Portellano-Ortiz and Conde-Sala, 2018; Prince et al., 1999).

Finally, regarding the fourth and the last aim of the study, our results showed adequate nomological validity of the EURO-D. Both factors, Affective Suffering and Lack of Motivation, were related in the expected directions as they were positively related to Loneliness and taking drugs for anxiety or depression, and negatively to Self-perceived Health and, findings that are in line with previous research (Erzen and Çikrikci, 2018; Salman and Lee, 2019).

5. Limitations and future research

Our study offers a comprehensive examination of the properties of the EURO-D scale, however, is not without limitations. Due to its cross-sectional design, one aspect that could be consider in further research may be to study the longitudinal stability of the EURO-D structure, since it is known that the symptomatology of depression tends to become chronic in old age (De la Torre-Luque et al., 2019).

Other aspect that could be improved is that our study only includes old people from Europe and Israel, it would be advisable to study in future research the invariance between countries and in other age groups. Cross-cultural comparative research of depressive symptoms requires that the invariance of the scale was proved in order to avoid cultural bias or other methodological artefacts (Maskileyson et al., 2021), the invariance across European countries of the EURO-D was proved (Castro-Costa et al., 2007; Maskileyson et al., 2021; Prince et al., 1999) and in some Latin America countries and Indian (Guerra et al., 2015), however, in other countries as China and Nigeria the properties

of the scale remain unclear (Guerra et al., 2015).

6. Conclusions

Confirmatory factorial analysis identified two factors: Affective Suffering, composed by the items related to depressed mood, suicidality, guilt, sleep, irritability, fatigue and tearfulness and Lack of Motivation, composed by pessimism, interest, appetite, concentration and enjoyment. Reliability in both factors was adequate and both factors have a positive relation with loneliness and negative with self-perceived health, as expected. In addition, the items don't show differential functioning by gender, but women have higher scores than men on both factors specially on Affective Suffering. In summary, the EURO-D scale has adequate psychometric properties to screening depression symptoms in old age and finding real differences between men and women.

Credit authorship contribution statement

Irene Fernández: Conceptualization & data analysis. Zaira Torres: Conceptualization, Writing – original draft. Sara Enrique: Writing - review & editing. Amparo Oliver: Writing - review & editing. José M. Tomás: Conceptualization– review & data analysis.

Conflict of Interest

The authors declare no conflict of interest.

References

Akosile, C.O., Mgbeojedo, U.G., Maruf, F.A., Okoye, E.C., Umeonwuka, I.C., Ogunniyi, A., 2018. Depression, functional disability and quality of life among Nigerian older adults: prevalences and relationships. Arch. Gerontol. Geriatr. 74, 39–43. https://doi.org/10.1016/j.archger.2017.08.011.

Alexopoulos, G.S., 2005. Depression in the elderly. Lancet 365 (9475), 1961–1970. https://doi.org/10.1016/S0140-6736(05)66665-2.

Asberg, M., Schalling, D., 1979. Construction of a new psychiatric rating instrument, the Comprehensive Psychopathological Rating Scale (CPRS). Prog. Neuropsychopharmacol. 3 (4), 405–412. https://doi.org/10.1016/0364-7722(79)90055-9.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4 (6), 561–571.

Bethmann, A., Bergmann, M., Scherpenzeel, A., 2019. SHARE Sampling Guide – Wave 8. Working Paper Series 33-2019.

Börsch-Supan, A., 2022. Survey of Health, Ageing And Retirement in Europe (SHARE) Wave 8. Release version: 8.0.0. SHARE-ERIC. https://doi.org/10.6103/SHARE. w8.800. Data set.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S., 2013. Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). Int. J. Epidemiol. 42, 992–1001. https:// doi.org/10.1093/ije/dyt088.

Brailean, A., Guerra, M., Chua, K.C., Prince, M., Prina, M.A., 2015. A multiple indicators multiple causes model of late-life depression in Latin American countries. J. Affect. Disord. 184, 129–136. https://doi.org/10.1016/j.jad.2015.05.053.

Castro-Costa, E., Dewey, M., Stewart, R., Banerjee, S., Huppert, F., Mendonca-Lima, C., Prince, M., 2007. Prevalence of depressive symptoms and syndromes in later life in ten European countries: the SHARE study. Br. J. Psychiatry 191 (5), 393–401. https://doi.org/10.1192/bjp.bp.107.036772.

Castro-Costa, E., Dewey, M., Stewart, R., Banerjee, S., Huppert, F., Mendonca-Lima, C., Prince, M., 2008. Ascertaining late-life depressive symptoms in Europe: an evaluation of the survey version of the EURO-D scale in 10 nations. The SHARE project. Int. J. Methods Psychiatr. Res. 17 (1), 12–29. https://doi.org/10.1002/ mpr.236.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model. 9, 233–255. https://doi.org/10.1207/ S15328007SEM0902_5.

Cohen, J., 1992. A power primer. Psychol. Bull. 112 (1), 155–159. https://doi.org/ 10.1037/0033-2909.112.1.155.

Copeland, J.R.M., Kelleher, M.J., Kellett, J.M., Gourlay, A.J., Gurland, B.J., Fleiss, J.L., Sharpe, L., 1976. A semi-structured clinical interview for the assessment of diagnosis and mental state in the elderly: the geriatric mental state schedule: I.Development and reliability. Psychol. Med. 6 (3), 439–449. https://doi.org/10.1017/ s0033291700015889.

Courtin, E., Knapp, M., Grundy, E., Avendano-Pabon, M., 2015. Are different measures of depressive symptoms in old age comparable? An analysis of the CES-D and Euro-D scales in 13 countries. Int. J. Methods Psychiatr. Res. 24 (4), 287–304. https://doi. org/10.1002/mpr.1489.

De la Torre-Luque, A., de la Fuente, J., Sanchez-Niubo, A., Caballero, F.F., Prina, M., Muniz-Terrera, G., Haro, J.M., Ayuso-Mateos, J.L., 2019. Stability of clinically relevant depression symptoms in old-age across 11 cohorts: a multi-state study. Acta Psychiatr. Scand. 140 (6), 541–551. https://doi.org/10.1111/acps.13107.

Erzen, E., Çikrikci, Ö., 2018. The effect of loneliness on depression: a meta-analysis. Int. J. Soc. Psychiatry 64 (5), 427–435. https://doi.org/10.1177/0020764018776349.

Finney, S.J., DiStefano, C., 2006. Non-normal and categorical data in structural equation modeling. In: Hancock, G.R., Mueller, R.O. (Eds.), Structural Equation Modeling: A Second Course. Information Age, Greenwich, CT, pp. 269–314.

Guerra, M., Ferri, C., Llibre, J., Prina, A.M., Prince, M., 2015. Psychometric properties of EURO-D, a geriatric depression scale: a cross-cultural validation study. BMC Psychiatry 15 (1), 1–14. https://doi.org/10.1186/s12888-015-0390-4.

Gurland, B., Golden, R.R., Teresi, J.A., Challop, J., 1984. The SHORT-CARE: an efficient instrument for the assessment of depression, dementia and disability. J. Gerontol. 39 (2), 166–169. https://doi.org/10.1093/geronj/39.2.166.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23 (1), 56. https://doi.org/10.1136/jnnp.23.1.56.

Hu, L., Bentler, P.M., 1999. Cut-off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55. https://doi.org/10.1080/10705519909540118.

Hughes, M.E., Waite, L.J., Hawkley, L.C., Cacioppo, J.T., 2004. A short scale for measuring loneliness in large surveys results from two population-based studies. Res. Aging 26, 655–672. https://doi.org/10.1177/0164027504268574.

Institute for Health Metrics and Evaluation, 2021. Global health data exchange. http ://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d78 0dffbe8a381b25e1416884959e88b.

Jirapramukpitak, T., Darawuttimaprakorn, N., Punpuing, S., Abas, M., 2009. Validation and factor structure of the Thai version of the EURO-D scale for depression among older psychiatric patients. Aging Ment.Health 13 (6), 899–904. https://doi.org/ 10.1080/13607860903046479.

Jodoin, M.G., Gierl, M.J., 2001. Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. Appl. Meas. Educ. 14, 329–349. https://doi.org/10.1207/S15324818AME1404_2.

Larraga, L., Saz, P., Dewey, M.E., Marcos, G., Lobo, A., 2006. Validation of the Spanish version of the EURO-D scale: an instrument for detecting depression in older people. Int.J.Geriatr.Psychiatry 21 (12), 1199–1205. https://doi.org/10.1002/gps.1642.

Magis, D., Beland, S., Tuerlinckx, F., De Boeck, P., 2010. A general framework and an R package for the detection of dichotomous differential item functioning. Behav. Res. Methods 42, 847–862. https://doi.org/10.3758/BRM.42.3.847.

Markush, R.E., Favero, R.V., 1973. Epidemiologic assessment of stressful life events, depressed mood, and psychophysiological symptoms: a preliminary report. In: Dohrenwend, B.S., Dohrenwend, B.P. (Eds.), Stressful Life Events: Their Nature And Effects. John Wiley & Sons.

Maskileyson, D., Seddig, D., Davidov, E., 2021. The EURO-D measure of depressive symptoms in the aging population: comparability across European countries and Israel. Front.Polit.Sci. 3, 1–12. https://doi.org/10.3389/fpos.2021.665004.

Millsap, R.E., 2012. Statistical Approaches to Measurement Invariance. Routledge. https://doi.org/10.4324/9780203821961.

Muthén, L.K., Muthén, B.O., 1998-2017. Mplus User's Guide, 8th ed. Muthén & Muthén, Los Angeles.

Narayanan, P., Swaminathan, H., 1996. Identification of items that show non-uniform DIF. Appl. Psychol. Meas. 20, 257–274. https://doi.org/10.1177/ 014662169602000306.

Portellano-Ortiz, C., Conde-Sala, J.L., 2018. Cognition and its association with the factors of the EURO-D: Suffering and Motivation. Findings from SHARE Wave 6. Int.J. Geriatr.Psychiatry 33 (12), 1645–1653. https://doi.org/10.1002/gps.4967.

Portellano-Ortiz, C., Garre-Olmo, J., Calvó-Perxas, L., Conde-Sala, J.L., 2018. Factor structure of depressive symptoms using the EURO-D scale in the over-50s in Europe. Findings from the SHARE project. Aging Ment. Health 22 (11), 1477–1485. https:// doi.org/10.1080/13607863.2017.1370688.

Prajwal, V., Arun, V., Ramya, M.C., Nagaraj, S., Krishnaveni, G.V., Kumaran, K., Krishna, M., 2021. Validation of EURO-D, a geriatric depression scale in South India: findings from the Mysore study of Natal effects on Ageing and Health (MYNAH). J. Affect. Disord. 295, 939–945. https://doi.org/10.1016/j.jad.2021.08.045.

Prince, M.J., Reischies, F., Beekman, A.T., Fuhrer, R., Jonker, C., Kivela, S.L., Copeland, J.R., 1999. Development of the EURO–D scale–a European Union initiative to compare symptoms of depression in 14 European countries. Br. J. Psychiatry 174 (4), 330–338. https://doi.org/10.1192/bjp.174.4.330.

R Core Team, 2021. R: A Language And Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1 (3), 385–401. https://doi.org/10.1177/ 014662167700100306.

Raykov, T., 2004. Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. Behav.Ther. 35, 299–331. https://doi.org/10.1016/ S0005-7894(04)80041-8.

Russel, D., Peplau, L.A., Cutrona, C.E., 1980. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J. Pers. Soc. Psychol. 39, 472–480. https://doi.org/10.1037/0022-3514.39.3.472.

Salman, A., Lee, Y.H., 2019. Spiritual practices and effects of spiritual well-being and depression on elders' self-perceived health. Appl. Nurs. Res. 48, 68–74. https://doi. org/10.1016/j.apnr.2019.05.018.

Schmitz, A., Lazarević, P., 2020. The gender health gap in Europe's ageing societies: universal findings across countries and age groups? Eur. J. Ageing 17 (4), 509. https://doi.org/10.1007/s10433-020-00559-6.

Sheikh, J.I., Yesavage, J.A., 1986. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. In: Brink, T.L. (Ed.), Clinical Gerontologist: A Guide to Assessment And Intervention. The Haworth Press Inc, pp. 165–173.

Svanborg, P., Åsberg, M., 1994. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr. Scand. 89 (1), 21–28. https://doi.org/10.1111/j.1600-0447.1994.tb01480.x.

Van de Velde, S., Bracke, P., Levecque, K., 2010. Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. Soc. Sci. Med. 71 (2), 305–313. https://doi.org/10.1016/j.socscimed.2010.03.035.

Ware, J.E., Gandek, B., 1998. Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. J. Clin. Epidemiol. 51, 903–912. https:// doi.org/10.1016/s0895-4356(98)00081-x.

World Health Organization, 2012. fact sheet No.369-Depression. Accessed on the 13th of October 2021 at. https://www.who.int/news-room/fact-sheets/detail/depression.

Wu, J.J., Wang, H.X., Yao, W., Yan, Z., Pei, J.J., 2020. Late-life depression and the risk of dementia in 14 countries: a 10-year follow-up study from the Survey of Health, Ageing and Retirement in Europe. J. Affect. Disord. 274, 671–677. https://doi.org/ 10.1016/j.jad.2020.05.059.

Zumbo, B.D., Thomas, D.R., 1997. A Measure of Effect Size for a Model-based Approach for Studying DIF. University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science, Prince George, Canada.

Zung, W.W., 1965. A self-rating depression scale. Archives of general psychiatry 12 (1), 63–70.