Validation of the Eight-Item Center for Epidemiologic Studies Depression Scale (CES-D) Among Older Adults

Jahanvash Karim, Robert Weisz, Zainab Bibi & Shafiq ur Rehman

Current Psychology
A Journal for Diverse Perspectives on Diverse Psychological Issues

ISSN 1046-1310

Curr Psychol
DOI 10.1007/s12144-014-9281-y

Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Author's personal copy

Curr Psychol
DOI 10.1007/s12144-014-9281-y

Validation of the Eight-Item Center for Epidemiologic Studies Depression Scale (CES-D) Among Older Adults

Jahanvash Karim · Robert Weisz · Zainab Bibi · Shafiq ur Rehman

© Springer Science+Business Media New York 2014

Abstract The aim of the present study was to examine the factor structure and psychometric properties of the 8-item short version of the Center for Epidemiologic Studies-Depression Scale (CES-D) in a large sample of European older adults. Data from The European Social Survey (ESS Round 6 Edition 1.1) provided the basis for this study (5,774 Males and 7,258 Females). Exploratory and confirmatory factor analyses provided support for a single factor structure. The results from multi-group confirmatory factor analyses revealed that the factor structure of the CES-D 8 is invariant across sexes, including invariance of item intercepts, item residuals, and item factor loadings. Moreover, the results provided support for the nomological validity of the scale. These results suggest that the shorter 8-item CES-D scale is a valid and reliable instrument of depression and extends the list of available instruments for screening depression among older adults.

Keywords Depression · CES-D 8 · Reliability · Factorial validity · Nomological validity

Depression, "a mental disorder that presents with depressed mood, loss of interest or pleasure, decreased energy, feelings of guilt or low self-worth, disturbed sleep or appetite, and poor concentration" (Marcus et al. 2012, p. 8), is a common health problem in older adults (Blazer 2009; Djernes 2006). The prevalence estimates of depression in elderly range from 10 to 15 % (Aisling et al. 2000; Kay et al. 1985; Liu

J. Karim (✉) · Z. Bibi
Institute of Management Sciences, University of Balochistan, Quetta, Pakistan
e-mail: j_vash@hotmail.com

Z. Bibi
e-mail: znb2005@hotmail.com

R. Weisz
CERGAM, IAE d'Aix-en-Provence, Université Paul Cézane, Marseille, France
e-mail: Robert.weisz@iae-aix.com

S. ur Rehman
Quaid-e-Azam University, Islamabad, Pakistan
e-mail: shafiq64@yahoo.com

Published online: 22 October 2014

Author's personal copy

et al. 1997; Newman et al. 1998; Schoevers et al. 2000; World Health Organization 2001). The benefits from an improved understanding of depression among older adults are likely to be substantial. Depression in older adults has implications for many life outcomes, including disability and impaired well-being (Beekman et al. 1997), suicide (Conwell and Brent 1995), alcohol and substance abuse (Koenig and Blazer 1996), morbidity and mortality (Gotlib and Hammen 2009), and general health (Royal College of Psychiatrists Working Group of the Faculty of Old Age Psychiatry 2005). Early detection, by administering time-effective and psychometrically sound screening instruments, can help in minimizing the adverse effects of depression. However, given the high prevalence of depression among elderly, it is worth mentioning that "healthy, normally functioning older adults are at no greater risk for depression than younger adults. What seem to be age-related effects on depression are attributable to physical health problems and related disability" (Roberts et al. 1997. p. 1384).

Given the strong and growing interest in depression, many psychometrically sound interviewer-rated and pencil-and-paper scales have been developed for screening depression in individuals. Compared to interviewer-rated depression scales, pencil-and-paper depression screening instruments have remained the focus of much attention, as they are often developed to screen depression quickly and inexpensively (e.g., the Zung Self Rated Rating Scale: Zung 1965; the Geriatric Depression Scale: Brink et al. 1982; the Beck Depression Inventory: Beck et al. 1997; and the Center for Epidemiological Studies Depression Scale: Radloff 1977). Most of these instruments were developed for general populations and their use with older adults calls for empirical validation.

The Center for Epidemiologic Studies–Depression Scale (CES–D) is a commonly used self-rating scale designed to measure depressive symptomatology (interpersonal relations, positive affect, depressed affect, and somatic activity) in clinical and non-clinical settings (Radloff 1977). The original CES-D scale consisted of 20 items. Although the CES-D 20-item version appears to be a highly reliable and valid instrument that has been widely used to screen depression in older adults (Beekman et al. 1997; Mui et al. 2001; Papassotiropoulos and Heun 1999), the administration of the CES-D can be problematic among older adults (Irwin et al. 1999). It may be difficult for older adults to remain focused for the length of time required to complete the instrument. Similarly, if the CES-D is employed in telephone interview, the number of items may decrease the participants' motivation in responding to the items. Given the problems associated with administration of the original 20-item scale, researchers have made efforts to develop shortened versions of the scale (Andresen et al. 1994; Kohout et al. 1993; Melchior et al. 1993; Santor and Coyne 1997; Van de Velde et al. 2009; 2008).

The CES-D 8-item is a commonly used abbreviated version of the CES-D 20 (Turvey et al. 1999). It is often preferred over other versions (e.g., CES-D 20 or CES-D 10) because of following reasons. First, the CES-D 8 is one of the most commonly used measures in many large sample-population based surveys such as the Asset and Health Dynamics among the Oldest Old (AHEAD) (Fonda and Herzog 2001), The Health and Retirement Study (HRS) (Steffick 2000), the European Social Survey (http://www.europeansocialsurvey.org), English Longitudinal Study of Ageing (ELSA) (Steptoe et al. 2013), Established Populations for Epidemiologic Studies of the Elderly (EPESE) (Cornoni-Huntley et al. 1990). Second, it has been widely used for screening depression among elderly (Turvey et al. 1999). Third, its internal consistency,

Author's personal copy

reliability and validity are almost identical to those of the 10-item CES-D and 20-item CES-D (O'Halloran et al. 2014).

This one-dimensional 8-item version was first embedded in The European Social Survey (ESS3 2006/07). Van de Velde et al. (2009) analyzed data from the ESS3 and found that a one-factor solution provided the best fit of the data to the model. In addition, their results provided support for the factorial invariance of the CES-D 8 across sexes and countries. Whilst initial studies (Van de Velde et al. 2009; 2008) have demonstrated acceptable psychometric properties of the scale in general population, psychometric issues of the scale need to be addressed in older adults.

The aim of the present study is to extend the utility of the CES-D 8 by assessing its psychometric qualities in a sample of older adults, where an early detection of depression is called for (Blazer 2009). To our knowledge this is the first study on this issue. First, we evaluated the unidimensional factor structure of the scale in a sample of elderly adults using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Second, we examined measurement invariance of the scale across sexes by the multi-group confirmatory factor analyses (Bollen 1989). We selected gender (for assessing measurement invariance) since depression scales are often used to compare depression scores across sexes (e.g., Kessler et al. 1993; Piccinelli and Wilkinson 2000; Weissman and Klerman 1977; Weissman et al. 1984). If valid comparisons are to be made across sexes then it is imperative that the psychometric properties of the scale are invariant across males and females. Measurement equivalence or measurement invariance is a statistical property of measurement that concerns the extent to which the same construct is being measured in same way across groups or item responses preserve their meaning across groups (Vandenberg and Lance 2000). "Comparing groups using a non-invariant measure is worthless or, put differently, is analogous to comparing apples and oranges" (Vlachopoulos et al. 2013, p.624). Once measurement equivalence is established, it is safe to conclude that differences between groups are meaningful (the groups differ with regards to the construct being measured and the differences are not due to differences in perceptions of meaning of the items across groups). Third, we assessed the known group validity of the scale by examining the mean differences between male and female subgroups. "Known-groups validity is demonstrated when individuals from populations that are 'known' to be 'high' (or 'low') on the construct being assessed score higher (or lower) on the purported measure of that construct than individuals 'known' to be 'low' (or 'high') on that construct" (Wallston 2005, p.624). Various studies have well-documented that females tend to score higher than males on depression scales (e.g., Kessler et al. 1993; Piccinelli and Wilkinson 2000; Weissman and Klerman 1977; Weissman et al. 1984).

Finally, we assessed the nomological validity of the CES-D 8. Nomological validity refers to a measure's ability to exhibit correlations (positively or negatively) with measures of other constructs in accordance with some established theory (Cronbach and Meehl 1955). Various studies have indicated that depression is associated with social trust (Alsund et al. 2010), autonomy (Fairbrother and Moretti 1998), optimism (Chang and Sanna 2001; Hart et al. 2008), self-esteem (Joiner et al. 1999; Orth and Robins 2013), subjective well-being (Cheung and Bagley 1998; Lewinsohn et al. 1991), social relationships/social isolation (Adams et al. 2004; Biordi and Nicholson 2009), and anxiety (Angst 1997). To demonstrate nomological validity of the scale, the composite CES-D 8 scores should have significant negative correlations with social

Author's personal copy

trust, autonomy, optimism, self-esteem, social relationships, satisfaction with life, and happiness, and a significant positive correlation with anxiety.

Method

Participants

The European Social Survey (ESS) is a large cross-national survey project that measures the attitudes, beliefs, and behaviors of people living in Europe. The ESS started in 2002 by the European Science Foundation (http://www.esf.org/). Every two years (2002, 2004, 2006, 2008, 2010, 2012), the survey is repeated in an increasing number of countries. Individuals for the survey are selected by random probability methods. The ESS questionnaires consist of two main parts, core module (repeated in each round) and rotating modules (modules on specific themes, for example, welfare attitudes and trust in justice).

The ESS is a publicly-available data set that can be utilized by researchers, social scientists and policy makers in a variety of ways. The major objective of the ESS is to study changes in economic, political, social, and cultural structures and processes across European countries (Jowell 2004). Data from the 2012 wave provide the basis for this study (ESS Round 6 Edition 1.1), particularly the ESS Personal and Social Well-Being Module which includes CES-D 8 and other well-being questions. The data collection of the sixth wave of the ESS took place in 2011/2012 in 24 countries (Belgium, Bulgaria, Switzerland, Cyprus, Czech Republic, Germany, Denmark, Estonia, Hungary, Spain, United Kingdom, Ireland, Israel, Iceland, Netherlands, Norway, Poland, Portugal, Russian Federation, Sweden, Slovenia, Slovakia, Kosovo) (N= 46,257). Due to large number of missing values on the CES-D 8 items, the data for Hungary was excluded from the analyses. Participants falling in the age bracket of 60 and older were selected for analyses. The final sample comprised 13,032 individuals (range =60 to 90 years; 5,774 Males; M_{age}=70.01, SD_{age}=7.40 and 7,258 Females; M_{age} =70.60, SD_{age}=7.79).

Measures

In the present study, analyses utilized both single-item and multi-item scales included in the ESS round 6. The variables of primary interest were CES-D 8, social trust, autonomy, optimism, self-esteem, subjective well-being, social relationships, and anxiety.

CES-D 8 In this study, we used the eight-item version of the original CES-D included in the ESS and suggested by Van de Velde et al. (2009). Item responses were rated on a four-point scale from (1) rarely or none of the time to (4) most or all of the time. To emphasize current state, the directions read "How often you have felt this way during the past week….". Eight items were "you felt depressed", "you felt everything you did was an effort", "your sleep was restless", "you were happy", "you felt lonely", "you enjoyed life", "you felt sad", and "you could not get going".

Author's personal copy

Subjective Well-Being There are two single-item variables measuring subjective well-being in the ESS: *satisfaction with life* ("How satisfied with life as a whole?"; rated on a ten-point scale from 0 [extremely dissatisfied] to10 [extremely satisfied])) and *happiness* ("How happy are you?"; rated on a ten-point scale from 0[extremely unhappy] to10[extremely happy]). Both single-item scales have demonstrated adequate reliability and validity (e.g., Abdel-Khalek 2006; Lucas and Donnellan 2012). Moreover, these single-item scales have been widely used in various large scale surveys, such as, German Socio-Economic Panel Study (GSOEP), The British Household Panel Study (BHPS), The Household, Income, and Labour Dynamics in Australia Study (HILDA), The Swiss Household Panel Study (SHP), and the World Values Survey (WVS).

Subjective Health A single item was used to assess the subjective health of participants: "How is your health in general?" Item responses were rated on a five-point scale from (1) very good to (5) very bad. The item was reverse-coded to facilitate analyses. This single-item measure has demonstrated acceptable reliability and validity (see Van Ginneken and Groenewold 2012).

Social Trust The ESS asks three questions about social trust: (1) "Most people can be trusted or you can't be too careful" (coded on a metric from 0 ["You can't be too careful"] to10 ["Most people can be trusted"]); (2) "Most people try to take advantage of you, or try to be fair" (coded on a metric from 0 ["Most people try to take advantage of me"] to10 ["Most people try to be fair"]) and; (3) "Most of the time people helpful or mostly looking out for themselves" (coded on a metric from 0 ["People mostly look out for themselves"] to10 ["People mostly try to be helpful"]). This scale, for measuring generalized social trust, has been used in previous large scale surveys (e.g. World Values Survey, European Values Survey, and General Social Survey).

Optimism A single-item measure of optimism was used in the present analyses: "I am always optimistic about my future". The item was measured with a five-point scale from 1="agree strongly" to 5="disagree strongly". The item was reverse-coded so that a higher score reflects higher optimism. This item was adopted from the Life Orientation Test (Scheir et al. 1994).

Self-Esteem Two items were used to measure self-esteem: "In general I feel very positive about myself" and "at times I feel as if I am a failure" with a 5-point scale ranging from 1="Agree strongly" to 5="Disagree strongly". First item was reverse-coded. Analyses utilized composite scores of two items so that a higher score reflects higher level of self-esteem. These items were adapted from the Rosenberg (1965) self-esteem scale (original items were "I take a positive attitude toward myself" and "I certainly feel useless at times").

Anxiety A single-item measure of anxiety was used in the present analyses: "Felt anxious, how often past week" (coded 1="None or almost none of the time", 2="Some of the time", 3="Most of the time", 4="All or almost all of the time"). This item was adopted from Norman et al. (2006) Overall Anxiety Severity and Impairment Scale (OASIS).

Author's personal copy

Autonomy A single-item measure of autonomy was used in the present analyses: "Free to decide how to live my life" with a 5-point scale ranging from 1="Agree strongly" to 5="Disagree strongly". The item was reverse-coded so that a higher score reflects higher level of autonomy. The item was adopted from the 21-item Basic Psychological Needs Scale (Ilardi et al. 1993). The item intends to assess the psychological needs for autonomy.

Results

Exploratory Factor Analysis (EFA)

Six thousand five hundred fifty seven participants were randomly selected from the overall data set (2,893 Males and 3,664 Females). The principal component method with varimax rotation was chosen for factor extraction.

Prior to EFA, the factorability of the eight CES-D items was examined. The overall Kaiser–Meyer–Olkin (KMO; Kaiser 1970) coefficient of sampling adequacy showed that the correlation matrix was suitable for factor analysis (KMO= 0.87). In addition, respective KMO values for individual items were>0.82, which is well-above the acceptable limit of 0.50. The Bartlett's Test of Sphericity was significant (χ^2 (28) =17,763.28, $p<0.05$), indicating that the correlation matrix was not an identity matrix or that the correlation matrix was significant for factor analysis. Finally, the communalities for all items were above 0.30 (see Table 1), confirming that variance of the original values was fairly explained by some underlying factor.

In line with suggestions outlined by Gorsuch (1983), four different criterions were used to determine the number of factors to extract: (1) Kaiser or minimum eigenvalue greater than 1 criterion (K1) (Kaiser 1960); (2) Cattell (1966) scree test; (3) parallel analysis (Horn 1965) and; (4) Velicer (1976) minimum average partial method (MAP). MAP and parallel analyses were carried out through SPSS programs developed by O'Connor (2000). Results supported extraction of only one factor. The first factor accounted for 48.74 % of the total variation (eigenvalue =3.98). Factor loadings ranged from 0.58 to 0.77. The item-total correlations ranged from 0.61 to 0.76. The mean of the eight-item scale was 1.79, with a standard deviation of 0.56. The internal consistency reliability of the scale (Cronbach's alpha) was 0.84. EFA using the principal axis factoring method also provided support for the unidimensionality of the scale and similar pattern of factor loadings.

Confirmatory Factor Analysis (CFA)

CFA was employed, using remaining data (N=6,475; Males =2,881 and Females =3,594), to verify the unidimensionality of the CES-D 8 revealed by EFA. AMOS 16 statistical software (Arbuckle 2007), with maximum-likelihood estimation, was used to perform confirmatory analyses.

Four fit indices were used in the present study: (1) The Chi-Square (χ^2) statistics (Bollen 1989); (2) the Tucker-Lewis Index (TLI; Tucker and Lewis 1973); (3) the

Author's personal copy

Table 1 Means, standard deviations, inter-item correlations, factor loadings, and item–total correlations for the CESD-8

	M	SD	1	2	3	4	5	6	7	F	h^2	MSA	I-T
1. Felt depressed, how often past week	1.54	0.73	–							0.77	0.6	0.89	0.76
2. Felt everything did as effort, how often past week	1.82	0.87	0.50	–						0.68	0.47	0.89	0.69
3. Sleep was restless, how often past week	1.86	0.89	0.41	0.36	–					0.58	0.34	0.93	0.61
4. Were happy, how often past week (R)	2.17	0.86	0.43	0.33	0.28	–				0.66	0.44	0.81	0.68
5. Felt lonely, how often past week	1.51	0.80	0.45	0.37	0.31	0.36	–			0.67	0.46	0.91	0.67
6. Enjoyed life, how often past week (R)	2.16	0.89	0.41	0.34	0.27	0.61	0.35	–		0.67	0.45	0.82	0.69
7. Felt sad, how often past week	1.64	0.75	0.58	0.43	0.38	0.41	0.53	0.43	–	0.77	0.60	0.88	0.76
8. Could not get going, how often past week	1.67	0.78	0.48	0.49	0.37	0.33	0.41	0.37	0.51	0.71	0.51	0.90	0.71

(N=6,557)

All correlations were significant at 0.01 level

F Factor Loadings, h^2 communalities, *MSA* Measures of Sampling Adequacy, *I-T* Item-Total Correlations

Comparative Fit Index (CFI: Bentler 1990); and (4) the Root-Mean-Square Error of Approximation (RMSEA; Browne and Cudeck 1993). The Chi-Square (χ^2) statistics (Bollen 1989) indicates the "magnitude of discrepancy between sample and fitted covariance matrices" (Hu and Bentler 1999, p. 32). Insignificant χ^2 values indicate good fit (Barrett 2007). The TLI is a relative fit index which assesses the model by comparing the chi-square value of the model to the chi-square value of the null model (in null model all measured variables are uncorrelated), with values close to 1.00 being indicative of good fit. The CFI compares the hypothesized model (sample covariance matrix) with null model in which all correlations among variables are zero, and where a value of CFI ≥0.95 is recognised as indicative of good fit (Hu and Bentler 1999). The RMSEA answers the question "How well would the model, with unknown but optimally chosen parameter values, fit the population covariance matrix if it were available?" (Browne and Cudeck 1993, p. 137–138). RMSEA values less than 0.05 indicate good fit.

Initial CFA results indicated a poor fit to the data with χ^2/df=96.56, p<0.05; GFI=0.93; CFI=0.89; and RMSEA=0.12 (low=0.11, high=0.12). An examination of modification indices indicated a strong evidence of correlated residuals between two reversed-coded items 4 ("Were happy, how often past week") and 6 ("Enjoyed life, how often past week") (modification index =1,356.85), which was considered as a reasonable one (Van de Velde et al. 2009). Therefore, correlated errors were modeled between items 4 and 6. This modified model (with correlated errors between items 4 and 6) was refit to the

Author's personal copy

data and resulted in a significant improvement in the model fit with χ^2/df=23.59, $p<0.05$; GFI=0.98; CFI=0.976; and RMSEA=0.059 (low=0.054, high=0.064). Fig. 1 presents the final one-factor measurement model. All standardized factor loadings were significant at $p<0.01$. The alpha reliability coefficient for the scale was 0.84.

Multiple Group Confirmatory Factor Analyses

Multi-group analyses were employed to test whether the structure of the CES-D 8 was invariant across gender (older women and older men). Prior to invariance analysis, one-factor baseline model of the CES-D 8 was tested separately for each sub-sample. Once a one-factor baseline model was established within each sub-sample separately, we tested the invariance of this model across the two sub-samples in subsequent multi-group CFAs (see Appendix Table 5 for inter-item correlations matrices of each gender group).

In line with suggestions outlined by Byrne (2010) and Widaman and Reise (1997), invariance was tested on four levels of nested models: configural, weak, strong, and strict invariance models. Equality constraints were imposed on factor loadings, item intercepts, and item residuals in stepwise fashion (Meredith 1993). Configural invariance requires that the number of factors and the respective factor-loading pattern be the same across male and female groups. In this model, parameters are freely estimated for each group in multi-group analysis or no equality constraints are imposed on parameters—factor loadings, intercepts, and unique item variances. This model is often known as a multi-group representation of the baseline models (Byrne 2010). In weak measurement invariance model factor loadings are constrained to equality across groups. The hypothesis of concern is that latent variable is related to the items in same way across groups. In strong measurement invariance model, in addition to factor loadings, the item intercepts are constrained to equality across groups. This indicates

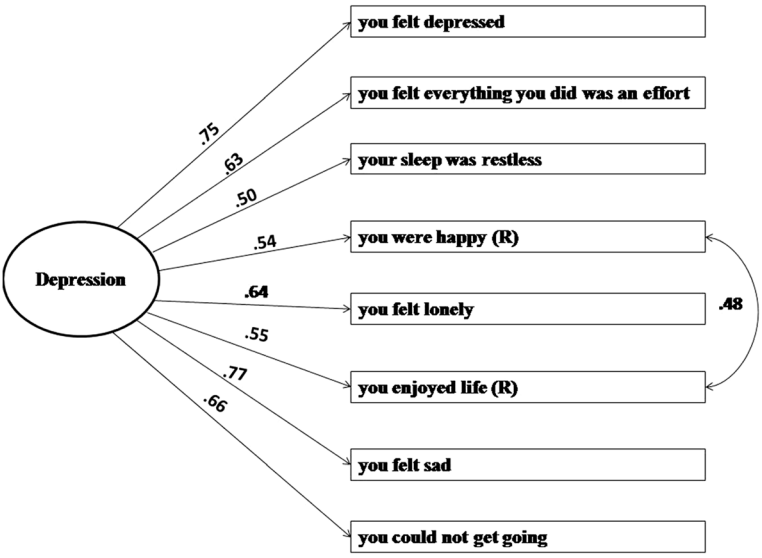

Fig. 1 Confirmatory factor analysis of the CES-D 8 (N=6,475)

Author's personal copy

that, in addition to invariant factor loadings, the item intercepts are invariant across groups. Finally, in strict measurement invariance model, in addition to factor loadings and item intercepts, item residuals (errors) are constrained to equality across groups. These nested models are compared to determine whether configural, weak, strong, and strict invariances are present across sub-samples. We used the $\Delta CFI<0.01$ criteria (Cheung and Rensvold 2002) for determining invariance of parameters (models) across sub-samples. If the more constrained model exhibits an adequate model fit and ΔCFI value (between the unconstrained and constrained models) exhibits that the constrained model does not lead to decrease in model fit as compared to the unconstrained model, the constrained parameters (factor loadings, intercepts, or error variances) are considered to be invariant across sub-samples. The $\Delta CFI<0.01$ would indicate invariance (Cheung and Rensvold 2002).

A one-factor congeneric measurement model fit well for both gender groups (see Table 2), and the item loadings were all significant ($p<0.001$) (see Table 3 for item loadings). Goodness-of-fit indices related to testing of unconstrained model (configural invariance model) suggested that the configural model represents the data well for both males and females (see Table 2 for fit indices and Table 3 for item loadings). Thus, configural invariance was supported, suggesting that males and females samples shared the same one-factor pattern. Next, to determine the invariance of item factor loadings (weak invariant model), equality constraints were imposed on item factor loadings. Fit indices indicated that this constrained model was acceptable (see Table 2). To determine the degree of weak invariance, we compared a metric invariant model (equality constraints imposed on item factor loadings) to the configural model (unconstrained model). The ΔCFI value between configural invariance model and weak invariance model was 0.001, indicating that item factor loadings were invariant across sexes. In addition to the item factor loadings, the indicator's intercepts were also constrained to be equivalent across sexes (strong invariant model). Multi-group analysis revealed that this strong invariant model was acceptable (see Table 2). Also, the ΔCFI value between

Table 2 Multi-group confirmatory factor analyses

	χ^2	DF	TLI	CFI	ΔCFI	RMSEA[CI]
Baseline models						
Male	216.73	19	0.960	0.973		0.060 (0.053, 0.067)
Female	261.37	19	0.965	0.976		0.060 (0.053, 0.066)
Invariance Models						
Configural (no equality constraints imposed)	478.10	38	0.963	0.975		0.042 (0.039, 0.046)
Weak (equality constraints imposed on λ's)	486.99	45	0.968	0.974	0.001	0.039 (0.036, 0.042)
Strong (equality constraints imposed on λ's and τ's)	668.12	53	0.962	0.965	0.009	0.042 (0.040, 0.045)
Strict (equality constraints imposed on λ's, τ's and θ's	725.08	61	0.965	0.962	0.003	0.038 (0.039, 0.046)

(N=6,475; Males =2,881 and Females =3,594). λ=item factor loadings, τ=item intercepts, θ=item residuals

Table 3 Parameter estimates of each model

	Baseline				Configural				Weak				Strong				Strict			
	Male		Female		Male		Female		Male		Female		Male		Female		Male		Female	
	Std	Unst	Std.	Unst	Std	Unst	Std	Unst	Std	Unst	Std	Unst	Std	Unst	Std	Unst	Std	Unst	Std	Unst
λ_{11}	0.73	1.00	0.75	1.00	0.73	1.00	0.75	1.00	0.73	1.00	0.75	1.00	0.74	1.00	0.76	1.00	0.72	1.00	0.77	1.00
λ_{21}	0.62	1.01	0.63	0.99	0.61	1.01	0.63	0.99	0.61	1.01	0.63	1.01	0.61	1.01	0.63	1.01	0.59	1.01	0.65	1.01
λ_{31}	0.48	0.82	0.49	0.77	0.48	0.82	0.49	0.77	0.47	0.79	0.50	0.79	0.48	0.80	0.51	0.80	0.47	0.80	0.52	0.80
λ_{41}	0.52	0.89	0.53	0.83	0.52	0.89	0.53	0.83	0.51	0.85	0.55	0.85	0.51	0.85	0.55	0.85	0.50	0.85	0.55	0.85
λ_{51}	0.60	0.88	0.65	0.94	0.60	0.88	0.65	0.94	0.61	0.92	0.64	0.92	0.62	0.92	0.64	0.92	0.60	0.92	0.66	0.92
λ_{61}	0.52	0.91	0.56	0.90	0.52	0.91	0.56	0.90	0.52	0.90	0.57	0.90	0.52	0.90	0.57	0.90	0.52	0.90	0.57	0.90
λ_{71}	0.75	1.05	0.77	1.04	0.75	1.05	0.77	1.04	0.75	1.04	0.77	1.04	0.75	1.05	0.78	1.05	0.74	1.05	0.78	1.05
λ_{81}	0.63	0.95	0.67	0.97	0.63	0.95	0.67	0.97	0.64	0.96	0.67	0.96	0.64	0.96	0.67	0.96	0.63	0.96	0.68	0.96
τ_{11}	1.41	1.41	1.59	1.59	1.41	1.41	1.59	1.59	1.41	1.41	1.59	1.59	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50
τ_{22}	1.71	1.71	1.87	1.87	1.71	1.71	1.87	1.87	1.71	1.71	1.87	1.87	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79
τ_{33}	1.74	1.74	1.95	1.95	1.74	1.74	1.95	1.95	1.74	1.74	1.95	1.95	1.85	1.85	1.85	1.85	1.85	1.85	1.85	1.85
τ_{44}	2.11	2.11	2.22	2.22	2.11	2.11	2.22	2.22	2.11	2.11	2.22	2.22	2.16	2.16	2.16	2.16	2.16	2.16	2.16	2.16
τ_{55}	1.41	1.41	1.58	1.58	1.41	1.41	1.58	1.58	1.41	1.41	1.58	1.58	1.49	1.49	1.49	1.49	1.49	1.49	1.49	1.49
τ_{66}	2.08	2.08	2.23	2.23	2.08	2.08	2.23	2.23	2.08	2.08	2.23	2.23	2.15	2.15	2.15	2.15	2.15	2.15	2.15	2.15
τ_{77}	1.51	1.51	1.72	1.72	1.51	1.51	1.72	1.72	1.51	1.51	1.72	1.72	1.61	1.61	1.61	1.61	1.62	1.62	1.62	1.62
τ_{88}	1.59	1.59	1.74	1.74	1.59	1.59	1.74	1.74	1.59	1.59	1.74	1.74	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67
θ_{11}	0.20	0.20	0.24	0.24	0.20	0.20	0.24	0.24	0.20	0.20	0.23	0.24	0.20	0.20	0.24	0.24	0.22	0.22	0.22	0.22
θ_{22}	0.41	0.41	0.47	0.47	0.41	0.41	0.47	0.47	0.41	0.41	0.47	0.47	0.41	0.41	0.47	0.47	0.45	0.45	0.45	0.45
θ_{33}	0.53	0.53	0.58	0.58	0.53	0.53	0.58	0.58	0.54	0.54	0.58	0.58	0.54	0.54	0.58	0.58	0.56	0.56	0.56	0.56
θ_{44}	0.51	0.51	0.54	0.54	0.51	0.51	0.54	0.54	0.51	0.51	0.54	0.54	0.52	0.52	0.54	0.54	0.53	0.53	0.53	0.53
θ_{55}	0.34	0.34	0.38	0.38	0.34	0.34	0.38	0.38	0.34	0.33	0.39	0.39	0.33	0.33	0.39	0.39	0.36	0.36	0.36	0.36

Author's personal copy

Table 3 (continued)

	Baseline				Configural				Weak				Strong				Strict			
	Male		Female		Male		Female		Male		Female		Male		Female		Male		Female	
θ_{66}	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54
θ_{77}	0.21	0.21	0.23	0.23	0.21	0.21	0.23	0.23	0.21	0.21	0.23	0.23	0.21	0.21	0.23	0.23	0.22	0.22	0.22	0.22
θ_{88}	0.33	0.33	0.36	0.36	0.33	0.33	0.36	0.36	0.33	0.33	0.36	0.36	0.33	0.33	0.36	0.36	0.35	0.35	0.35	0.35

Unst. Unstandardized estimates, *Std*. standardized estimates, λ item factor loadings, τ item intercepts, θ item residuals

Author's personal copy

Author's personal copy

strong invariance model and weak invariance model was less than 0.01, suggesting that, in addition to factor loadings, intercepts of each item were also invariant across gender groups. Finally, in addition to factor loadings and item intercepts, residual item variances for each item were constrained to be equal across the two gender groups (strict invariance model). The strict invariance model was acceptable (see Table 2). Also, the ΔCFI value between strict invariance model and strong invariance model was less than 0.01. In sum, results indicated that the item factor loadings (weak invariance), item intercepts (strong invariance), and item residual variances (strict invariance) were invariant across sexes. The parameter estimates of four nested models are presented in Table 3.

Significant differences in depression scores across sexes were observed, with older females (M=1.87, SD=0.58) scoring higher than older males (M =1.70, SD=0.52), $t_{(13030)}$=−17.27, p<0.01, d=0.30.

Regarding nomological validity, results revealed that depression measured by the CES-D 8 correlated significantly in negative direction with life satisfaction, happiness, social trust, self-esteem, optimism, subjective health, autonomy, and social relationships and correlated significantly in positive direction with anxiety (see Table 4).

Discussion

The CES-D 8 was developed to screen depression symptoms but has limited psychometric support for use with older adults. Psychometric researches indicate that factor structures of psychological measurements often differ among elderly compared with adults (Blazer 2009). The aim of the present study was to evaluate the psychometric properties of the 8-item short version of the CES-D and to test measurement invariance across sexes in a large sample of older adults.

The first objective of the present study was to examine, using EFA and CFA approach, the factorial structure of the CES-D 8 in a sample of older adults. The EFA using the principal component analysis and principal axis factoring method provided support for a single factor structure and a similar pattern of factor loadings. In the current study, we employed more rigorous and accurate factor extraction criteria (Kaiser's criterion, Cattell's scree test, HPA, and MPA). Therefore, it is recommended that attempting to extract more factors would be considered over factoring. Specifically, we found the internal consistency of the 8-item CES-D scale to be adequate. The CFA results indicated that the hypothesized single-factor provided a satisfactory fit to the data. These results are inline with previous findings (Van de Velde et al. 2009; 2008). However, the CES-D 8 factor structure showed better fit indices (particularly the RMSEA) when covariance between errors of item 4 ("Were happy, how often past week") and item 6 ("Enjoyed life, how often past week") were allowed. This suggests that there is a perceived redundancy between these items.

Additionally, multi-group CFAs were performed to assess the measurement invariance/equivalence of the 8-item CES-D across sexes. The results indicated that (i) a single-factor model was an acceptable description of the data for both males and females in older adults (configural invariance), (ii) latent construct (depression) was

Table 4 Correlations between CES-D 8 and other variables

	1	2	3	4	5	6	7	8	9	10	11	*M*	*SD*
1. CES-D 8		0.14**	−0.52**	−0.56**	−0.28**	−0.45**	−0.46**	−0.52**	−0.33**	−0.21**	0.59**	1.87	0.59
2. Age	0.10**		0.01	−0.02	0.05**	0.01	−0.08**	−0.20**	−0.03**	0.01	0.05**	70.60	7.80
3. Life satisfaction	−0.52**	0.04**		0.72**	0.38**	0.28**	0.40**	0.40**	0.28**	0.18**	−0.38**	6.90	2.45
4. Happiness	−0.57**	0.02	0.71**		0.35**	0.32**	0.42**	0.40**	0.29**	0.20**	−0.40**	7.05	2.21
5. Social trust	−0.29**	0.03**	0.38**	0.37**		0.09**	0.22**	0.24**	0.16**	0.17**	−0.23**	5.24	2.10
6. Self-esteem	−0.45**	0.03*	0.31**	0.35**	0.11**		0.43**	0.24**	0.27**	0.13**	−0.30**	3.70	0.77
7. Optimism	−0.43**	−0.03*	0.41**	0.41**	0.24**	0.45**		0.32**	0.33**	0.12**	−0.32**	3.61	0.96
8. Subjective health	−0.50**	−0.16**	0.36**	0.35**	0.23**	0.25**	0.30**		0.23**	0.16**	−0.33**	3.29	0.93
9. Autonomy	−0.34**	0.01	0.27**	0.28**	0.15**	0.32**	0.34**	0.22**		0.13**	−0.24**	4.03	0.89
10. Social relationships	−0.21**	−0.04**	0.16**	0.20**	0.13**	0.13**	0.12**	0.17**	0.12**		−0.16**	4.70	1.65
11. Anxiety	0.53**	0.03*	−0.35**	−0.36**	−0.21**	−0.28**	−0.28**	−0.30**	−0.23**	−0.12**		1.72	0.79
M	1.70	70.01	7.13	7.29	5.24	3.83	3.71	3.45	4.11	4.61	1.53		
SD	0.52	7.41	2.36	2.11	2.04	0.74	0.92	0.93	0.84	1.61	0.71		

Females are above diagonal, males are below diagonal. (N=6,475; Males =2,881 and Females =3,594)

* $p<0.05$ and * $p<0.01$

Author's personal copy

Author's personal copy

related to the eight items in a similar fashion across sexes (weak invariance or factor loadings non-invariance), (iii) the CES-D 8 items were scored systematically lower or higher across males and females irrespective of their level on the latent construct (strong invariance or intercepts non-invariance), and (iv) the items of the CES-D 8 measure the depression with more (or less) measurement error in the two subgroups (strict invariance or errors non-invariance) or the items are equally reliable across sexes.

The third objective of this study was to compare older males' and older females' scores on the CES-D 8. Older females scored higher than older males on the CES-D 8. These results confirm previous findings suggesting that women report higher levels of depression than man (e.g., Kessler et al. 1993; Piccinelli and Wilkinson 2000; Weissman and Klerman 1977; Weissman et al. 1984). Many factors contribute to the prevalence of higher depression rates in females. For example, compared to men, women are more likely to perceive a greater lack of respect and power in society, are more likely to face harassment, are more likely to have low income and live in poverty, are more likely to have poorer self-concepts, and are less able to handle stressful events effectively (see Keita 2007 and Nolen-Hoeksema 2001). All these factors ultimately lead to depression. Additional analyses revealed that compared to older men, older women scored higher on all items. These results were in line with previous findings that found that women, in comparison with men, appear to be less happy (Mrozek and Kolarz l998), express sadness with more intensity (Balswick and Avertt 1977), experience more sleep problems (Zhang and Wing 2006), and experience more loneliness (Page and Cole 1991).

The fourth objective of this study was to examine the nomological validity of the 8-item CES-D scale with measures of social trust, autonomy, optimism, self-esteem, satisfaction with life, happiness, social relationships, subjective health, and anxiety. The results revealed that the scores on the scale were significantly correlated with theoretically relevant variables, which is consistent with results from previous studies (Adams et al. 2004; Alsund et al. 2010; Biordi and Nicholson 2009; Cheung and Bagley 1998; Chang and Sanna 2001; Fairbrother and Moretti 1998; Hart et al. 2008; Joiner et al. 1999; Lewinsohn et al. 1991; Orth and Robins 2013) and supported the nomological validity of the CES-D 8. In addition, current research indicated a strong relationship between anxiety and CES-D 8, confirming that "anxiety and depression can be defined as emotional states or as clinical syndromes" (Madden et al. 2000, p. 277).

Although results for psychometric properties of the CES-D 8 are promising, there are some limitations to the present research. First, we did not determine sensitivity (ability of the CES-D 8 to correctly identify people experiencing depression), specificity (ability of the CES-D 8 to correctly identify people not experiencing depression), and receiver operating characteristics (ROC) curves (Murphy et al. 1987) for determining a cutoff value for the scale. It is recommended that future studies should employ ROC analysis for selecting the best threshold score for the CES-D 8 and compare the sensitivity and specificity of the CES-D 8 against a criterion diagnosis of depressive disorder (e.g., the Beck Depression Inventory: Beck et al. 1997) in elderly sample. Second, we evaluated the psychometric properties of the CES-D 8 in a nonclinical sample, and no incremental utility of the scale was found over other similar scales. Third, self-reported responses on the CES-D 8 and other variables could lead to

Author's personal copy

common method variance problems. Further validation of the scale should explore relationships with different depression measures (e.g., interviewer-rated depression rating scales) and test causal relationships with antecedents (e.g., self-esteem, self-efficacy, locus of control) and outcomes (e.g., satisfaction with life) using longitudinal or experimental designs. Fourth, as the ESS is a cross-sectional in nature (participants are not followed up with the same questionnaire), we were unable to assess the test-retest reliability and longitudinal invariance of the scale. Future studies should assess the test-retest reliability and test measurement invariance of the scale using a longitudinal framework. Fifth, although we have employed item-total correlations and multi-group confirmatory factor analysis procedures, future studies should employ more sophisticated differential item functioning (DIF) procedures for determining item bias. Sixth, this study relied exclusively on a sample of older adults from Europe (predominantly western cultures). Future studies should assess the psychometric properties of the scale in Eastern cultures to enhance its generalizability. Finally, the usage of single-item measurement of constructs in psychological research has been a subject of debate in literature. On the positive side, single-item scales are easy to administer, flexible, reduce response bias and are not monotonous to complete (Drolet and Morrison 2001; Gardner et al. 1998; Pomeroy et al. 2001). In contrast, various researchers strongly advocate the use of multi-item scales (see Boyd et al. 2005; Churchill 1979; Diamantopoulos and Winklhofer 2001; Viswanathan 2005). More importantly, compared to multi-item scales, single-item measures lack construct validity because they do not tap the construct from different dimensions (Baumgartner and Homburg 1996; Nunnally and Bernstein 1994; Wirtz and Lee 2003). Therefore, as one of the anonymous reviewers of this manuscript rightly pointed out, a stronger case for the nomological validity of the CESD-8 would be made if correlations were examined with well-established multi-item/multi-dimensional scales such as the Rosenberg's self-esteem scale (Rosenberg 1965), The Oxford Happiness Questionnaire (Hills and Argyle 2002), Life Orientation Test (Scheir et al. 1994), The Hamilton Anxiety Scale (Maier et al. 1988), and The index of autonomous functioning (Weinstein et al. 2012).

Conclusion

Given the problems associated with the use of self-report depression assessing instruments in the elderly (e.g., administration and scoring time) (Gallagher et al. 1980), a number of researchers have called for the development and validation of shortened versions of the depression scales for the elderly (McNair 1979; Montorio and Izal 1996). According to the results of this study, the 8-item CES-D is not only a valid instrument, it also has several advantages over longer 20-item version. It is easy to read, easy to understand, simple in format, extremely easy to administer, and requires minimum time. Based on the present results we suggest that the CES-D 8 is a suitable instrument for measuring or screening depression in older adults. Hence, the CES-D 8 opens up further possibilities for depression research in clinical or non-clinical settings, as its brief nature may lead to wide-scale adoption by researchers and clinical psychologists.

Author's personal copy

Appendix

Table 5 Item correlation matrices of each gender group

	M	SD	S	K	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
Male (*N=2,881*)											
Item 1	1.42	0.67	1.71	2.84							
Item 2	1.71	0.81	1.01	0.39	0.48						
Item 3	1.74	0.84	0.97	0.29	0.36	0.35					
Item 4	2.12	0.84	0.33	−0.55	0.40	0.29	0.26				
Item 5	1.41	0.73	1.89	3.13	0.42	0.31	0.24	0.35			
Item 6	2.08	0.86	0.37	−0.60	0.38	0.31	0.26	0.63	0.34		
Item 7	1.52	0.69	1.37	1.83	0.56	0.41	0.34	0.41	0.51	0.40	
Item 8	1.60	0.74	1.21	1.20	0.44	0.47	0.34	0.30	0.38	0.32	0.47
Female (*N=3,594*)											
Item 1	1.59	0.74	1.21	1.16							
Item 2	1.88	0.89	0.79	−0.15	0.51						
Item 3	1.95	0.88	0.71	−0.15	0.38	0.36					
Item 4	2.23	0.87	0.24	−0.60	0.43	0.30	0.26				
Item 5	1.59	0.82	1.36	1.16	0.47	0.39	0.30	0.37			
Item 6	2.23	0.89	0.21	−0.76	0.44	0.35	0.28	0.63	0.39		
Item 7	1.72	0.76	0.98	0.79	0.59	0.43	0.37	0.44	0.56	0.44	
Item 8	1.75	0.82	0.98	0.44	0.50	0.50	0.34	0.33	0.41	0.36	0.52

References

Abdel-Khalek, A. (2006). Measuring happiness with a single-item scale. *Social Behavior and Personality, 34*(2), 139–150.

Adams, K. B., Sanders, S., & Auth, E. A. (2004). Loneliness and depression in independent living retirement communities: Risk and resilience factors. *Aging and Mental Health, 8*, 475–485.

Aisling, D., Kirby, M., Bruce, I., Cunningham, C., Coakley, D., & Lawlor, B. A. (2000). Three-year prognosis of depression in the community-dwelling elderly. *British Journal of Psychiatry, 176*, 453–7.

Alsund, C., Starrin, B., & Nilsson, K. W. (2010). Social capital in relation to depression, musculoskeletal pain, and psychosomatic symptoms: A cross-sectional study of a large population-based cohort of Swedish adolescents. *BMC Public Health, 10*, 715.

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D. *American Journal of Preventive Medicine, 10*(2), 77–84.

Angst, J. (1997). Depression and anxiety: implications for nosology, course, and treatment. *Journal of Clinical Psychiatry, 58*(Suppl 8), 3–5.

Arbuckle, J. L. (2007). *Amos™ 16 user's guide*. Chicago: SPSS.

Balswick, J., & Avertt, C. P. (1977). Differences in expressiveness: Gender interpersonal orientation, and perceived parental expressiveness as contributing factors. *Journal of Marriage and the Family, 39*, 121–127.

Author's personal copy

Barrett, P. (2007). Structural equation modelling: Adjudging model Fit. *Personality and Individual Differences, 42*(5), 815–24.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing research: A review. *International Journal of Research in Marketing, 13*(2), 139–161.

Beck, A. T., Guth, D., Steer, R. A., & Ball, R. (1997). Screening for major depression disorders in medical inpatients with the beck depression inventory for primary care. *Behavior Research and Therapy, 35*, 785–791.

Beekman, A. T. F., Deeg, D. J. H., Braam, A. W., Smit, J. H., & Van Tilburg, W. (1997). Consequences of major and minor depression in later life: a study of disability, well-being and service utilization. *Psychological Medicine, 27*, 1397–1409.

Bentler, P. M. (1990). Comparative Fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–46.

Biordi, D. L., & Nicholson, N. R. (2009). Social isolation. In P. D. Larsen, F. W. Whitney, & I. M. Lubkin (Eds.), *Chronic illness: Impact and intervention* (pp. 85–115). Boston: Jones and Bartlett Publishers.

Blazer, D. G. (2009). Depression in late life: Review and commentary. *Focus, 7*(1), 118–136.

Bollen, K. A. (1989). *Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics*. New York: Wiley.

Boyd, B. K., Gove, S., & Hitt, M. A. (2005). Construct measurement in strategic management research: Illusion or reality? *Strategic Management Journal, 26*(3), 239–257.

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist, 1*(1), 37–43.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage.

Byrne, B. M. (2010). *Structural equation modeling with Amos: Basic concepts, applications, and programming* (2nd ed.). New York: Taylor and Francis Group.

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

Chang, E. C., & Sanna, L. J. (2001). Optimism, pessimism, and positive and negative affectivity in middle-aged adults: a test of a cognitive-affective model of psychological adjustment. *Psychology and Aging, 16*(3), 524–31.

Cheung, C.-K., & Bagley, C. (1998). Validating an American scale in Hong Kong: The center for epidemiological studies depression scale (CES-D). *The Journal of Psychology, 132*(2), 169–186.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*, 64–73.

Conwell, Y., & Brent, D. (1995). Suicide and aging. I: patterns of psychiatric diagnosis. *International Psychogeriatrics, 7*(2), 149–64.

Cornoni-Huntley, J., Blazer, D.G., Lafferty, M.E., Everett, D.F., Brock, D.B. & Farmer, M.E (Eds). (1990*). Established populations for epidemiologic studies of the elderly.* Volume II. Resource data book (NIH Publication 90–495). Bethesda, Md.: National Institute on Aging.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Vand de Velde, S., Bracke, P., & Levecque, K. (2008). The psychometric properties of the CES-D 8 depression inventory and the estimation of cross-national differences in the true prevalence of depression. International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC), 25–28 June, 2008, Berlin, Berlin-Brandenburg Academy of Sciences and Humanities). Retrieved from http://soc.kuleuven.be/ceso/dagvandesociologie/papers/CESD8%20in%20comparative%20perspective%20Vandevelde-Bracke-Levecque.pdf

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277.

Djernes, K. J. (2006). Prevalence and predictors of depression in populations of elderly: A review. *Acta Psychiatrica Scandinavica, 113*, 372–387.

Drolet, A. L., & Morrison, D. G. (2001). Rejoinder to grapentine. *Journal of Service Research, 4*(2), 155–160.

Fairbrother, N., & Moretti, M. (1998). Sociotropy, autonomy, and self-discrepancy: Status in depressed, remitted depressed, and control participants. *Cognitive Therapy and Research, 22*(3), 279–296.

Fonda, S. J., & Herzog, A. R. (2001). Patterns and risk factors of change in somatic and mood symptoms among older adults. *Annals of Epidemiology, 11*(6), 361–368.

Gallagher, D., Thompson, L. W., & Levy, S. M. (1980). Clinical psychological assessment of older adults. In L. W. Poon (Ed.), *Aging in the 1980s* (pp. 19–40). New York: American Psychological Association.

Author's personal copy

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement, 58*(6), 898–915.

Gorsuch, R. L. (1983). *Factor analysis (2nd ed.)*. Hillsdale: Erlbaum.

Gotlib, I. H., & Hammen, C. L. (2009). *Handbook of depression*. London: The Guilford Press.

Hart, S. L., Vella, L., & Mohr, D. C. (2008). Relationships among depressive symptoms, benefit-finding, optimism, and positive affect in multiple sclerosis patients after psychotherapy for depression. *Health Psychology, 27*(2), 230–8.

Hills, P., & Argyle, M. (2002). The Oxford happiness questionnaire : A compact scale for the measurement of psychological well-being. *Personality and Individual Differences, 33*(7), 1073–1082.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 32*, 179–185.

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), 1–55.

Ilardi, B. C., Leone, D., Kasser, R., & Ryan, R. M. (1993). Employee and supervisor ratings of motivation: Main effects and discrepancies associated with job satisfaction and adjustment in a factory setting. *Journal of Applied Social Psychology, 23*, 1789–1805.

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item center for epidemiological studies depression scale (CES-D). *Archives of Internal Medicine, 159*(15), 1701–1704.

Joiner, T. E., Katz, J., & Lew, A. (1999). Harbingers of depressotypic reassurance seeking: Negative life events, increased anxiety, and decreased self-esteem. *Personality and Social Psychology Bulletin, 25*, 632–639.

Jowell, R. (2004). *European Social Survey Round 1: 2002–2003 (Technical Report, ed. 2)*. London: National Centre for Social Research.

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

Kaiser, H. F. (1970). A second-generation Little Jiffy. *Psychometrika, 35*, 401–415.

Kay, D. W., Henderson, A. S., Scott, R., Wilson, J., Rickwood, D., & Grayson, D. A. (1985). Dementia and depression among the elderly living in the Hobart community: The effect of the diagnostic criteria on the prevalence rates. *Psychological Medicine, 15*, 771–88.

Keita, G. P. (2007). Psychosocial and cultural contributions to depression in women: Consideration for women midlife and beyond. *Journal of Managed Care Pharmacy, 13*(9), s12–s15. suppl s-a.

Kessler, R. C., Mcgonagle, K. A., Swartz, M., Blazer, D. G., & Nelson, C. B. (1993). Sex and depression in the national comorbidity survey.1: Lifetime prevalence, chronicity and recurrence. *Journal Affective Disorders, 29*(2–3), 85–96.

Koenig, H. G., & Blazer, D. G. (1996). Depression. In J. E. Birren (Ed.), *Encyclopedia of gerontology: Age, aging, and the aged* (Vol. I, pp. 415–428). San Diego: Academic.

Kohout, F. J., Berkman, L. F., Evans, D. A., & Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D depression symptoms index. *Journal of Aging and Health, 5*(2), 179–193.

Lewinsohn, P. M., Redner, J. E., & Selley, J. R. (1991). The relationship between life satisfaction and psychosocial variables: New perspectives. In F. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective well-being: An interdisciplinary perspective* (pp. 141–169). Oxford: Pergamon Press.

Liu, C. Y., Wang, S. J., Teng, E. L., Fuh, J. L., Lin, C. C., Lin, K. N., et al. (1997). Depressive disorders among older residents in a Chinese rural community. *Psychological Medicine, 27*, 943–9.

Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from National panel studies. *Social indicator Research, 105*(3), 323–331.

Madden, T. E., Barrett, L. F., & Pietromonaco, P. R. (2000). Sex differences in anxiety and depression: Empirical evidence and methodological questions. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 277–298). New York: Cambridge University Press.

Maier, W., Buller, R., Philipp, M., & Heuser, I. (1988). The Hamilton anxiety scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *Journal of Affective Disorders, 4*(1), 61–8.

Marcus, M., Yasamy, T. Y., Ommeren, M., Chisholm, D., & Saxena, S. (2012). Depression: A global health concern. In Depression: A global crisis (pp. 6–8).World Federation for Mental Health: Occoquan, VA. Retrieved from http://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf.

McNair, D. M. (1979). Self-rating scales for assessing psychopathology in the elderly. In A. Raskin & L. F. Jarvik (Eds.), *Psychiatric symptorns and cognitive loss in the elderly* (pp. 157–267). Washington: Wiley.

Author's personal copy

Melchior, L. A., Huba, G. J., Brown, V. B., & Reback, C. J. (1993). A short depression index for women. *Educational and Psychological Measurement, 53*(4), 1117–1125.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*, 525–542.

Montorio, I., & Izal, M. (1996). The geriatric depression scale: A review of its development and utility. *International Psychogeriatrics, 8*(1), 103–112.

Mrozek, D. K., & Kolarz, C. M. (l998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology, 75*(5), 1333–1349.

Mui, A., Burnett, D., & Chen, L. M. (2001). Cross-cultural assessment of geriatric depression: A review of the CES-D and the GDS. *Journal of Mental Health and Aging, 7*(1), 137–64.

Murphy, J. M., Berwick, D. M., Weinstein, M. C., Borus, J. F., Budman, S. H., & Klerman, G. L. (1987). Performance of screening and diagnostic tests: Application of receiver operating characteristic analysis. *Archives of General Psychiatry, 44*, 550–555.

Newman, S. C., Bland, R. C., & Orn, H. T. (1998). The prevalence of mental disorders in the elderly in Edmonton: A community survey using GMS-AGECAT. *Canadian Journal of Psychiatry, 43*, 910–4.

Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science, 10*(5), 173–175.

Norman, S. B., Cissel, S. H., Means-Christensen, A. J., & Stein, M. B. (2006). Development and validation of an overall anxiety severity and impairment scale (OASIS). *Depression and Anxiety, 23*(4), 245–9.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

O'Halloran, A. M., Kenny, R. A., & King-Kallimanis, B. L. (2014). The latent factors of depression from the short forms of the CES-D are consistent, reliable and valid in community-living older adults. *European Geriatric Medicine, 5*, 97–102.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396–402.

Orth, U., & Robins, R. W. (2013). Understanding the link between low self-esteem and depression. *Current Directions in Psychological Science, 20*(10), 1–6. doi:10.1177/0963721413492763.

Page, R. M., & Cole, G. (1991). Demographic predictors of self reported loneliness in adults. *Psychological Reports, 68*, 939–945.

Papassotiropoulos, A., & Heun, R. (1999). Screening for depression in the elderly: a study on misclassification by screening instruments and improvement of scale performance. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 23*(3), 431–446.

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression: Critical review. *British Journal of Psychiatry, 177*, 486–92.

Pomeroy, I. M., Clark, C. R., & Philip, I. (2001). The effectiveness of very short scales for depressions screening in elderly medical patients. *International Journal of Geriatric Psychiatry, 16*(3), 321–326.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurements, 3*, 385–401.

Roberts, K. E., Kaplan, G. A., Shema, S. J., & Strawbridge, W. J. (1997). Does growing old increase the risk for depression? *American Journal of Psychiatry, 154*(10), 1384–1390.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.

Royal College of Psychiatrists Working Group of the Faculty of Old Age Psychiatry. (2005). *Who cares wins: Improving the outcome for older people admitted to the general hospital*. London: Royal College of Psychiatrists.

Santor, D. A., & Coyne, J. C. (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment, 9*(3), 233–243.

Scheir, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*, 1063–1078.

Schoevers, R. A., Geerlings, M. I., Beekman, A. T., Penninx, B. W., Deeg, D. J., Jonker, C., et al. (2000). Association of depression and gender with mortality in old age. *British Journal of Psychiatry, 177*, 336–42.

Steffick, E. D. (2000). Documentation of affective functioning measures in the health and retirement study. Survey Research Center, University of Michigan, Ann Arbor, MI. Retrieved from http://hrsonline.isr.umich.edu/sitedocs/userg/dr-005.pdf, on Feburary 21, 2014.

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: the English longitudinal study of ageing. *International Journal of Epidemiology, 42*(6), 1640–1648.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

Author's personal copy

Turvey, C. L., Wallace, R. B., & Herzog, R. (1999). A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly. *International Psychogeriatrics, 11*(2), 139–148.

Van de Velde, S., Levecque, K., & Bracke, P. (2009). Measurement equivalence of the CES-D 8 in the general population in Belgium: a gender perspective. *Archives of Public Health, 67*, 15–29.

Van Ginneken, J. K., & Groenewold, G. (2012). A single-vs. multi-item self-rated health status measure: A 21-country study. *The Open Health Journal, 5*, 1–9.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–69.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321–327.

Viswanathan, M. (2005). *Measurement error and research design*. London: Sage Publications.

Vlachopoulos, S. P., Asci, F. H., Cid, L., Ersoz, G., Gonzalez-Cutre, D., Moreno-Murcia, J. A., & Moutao, J. (2013). Cross-cultural invariance of the basic psychological needs in exercise scale and need satisfaction latent mean differences among Greek, Spanish, Portuguese and Turkish samples. *Psychology of Sport and Exercise, 14*, 622–631.

Wallston, K. A. (2005). The validity of the multidimensional Health Locus of Control scales. *Journal of Health Psychology, 10*(4), 623–631.

Weinstein, N., Przybylski, A. K., & Ryan, R. M. (2012). The index of autonomous functioning: Development of a scale of human autonomy. *Journal of Research in Personality, 46*, 397–413.

Weissman, M. M., & Klerman, G. L. (1977). Sex-differences and epidemiology of depression. *Archives of General Psychiatry, 34*(1), 98–111.

Weissman, M. M., Leaf, P. J., Holzer, C. E., Myers, J. K., & Tischler, G. L. (1984). The epidemiology of depression: An update on sex-differences in rates. *Journal of Affective Disorders, 7*(3–4), 179–88.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington DC: American Psychological Association.

Wirtz, J., & Lee, M. C. (2003). An examination of the quality of context-specific applicability of commonly used customer satisfaction measures. *Journal of Service Research, 5*(4), 345–355.

World Health Organization. (2001). Mental Health: New understanding, New Hope. Retrieved from http://www.who.int/whr/2001/en/whr01_en.pdf on November, 2013.

Zhang, B., & Wing, Y.-K. (2006). Sex differences in insomnia: a meta- analysis. *Sleep, 29*, 85–93.

Zung, W. W. K. (1965). A. Self-rating depression scale. *Archives of General Psychiatry, 12*, 63–70.