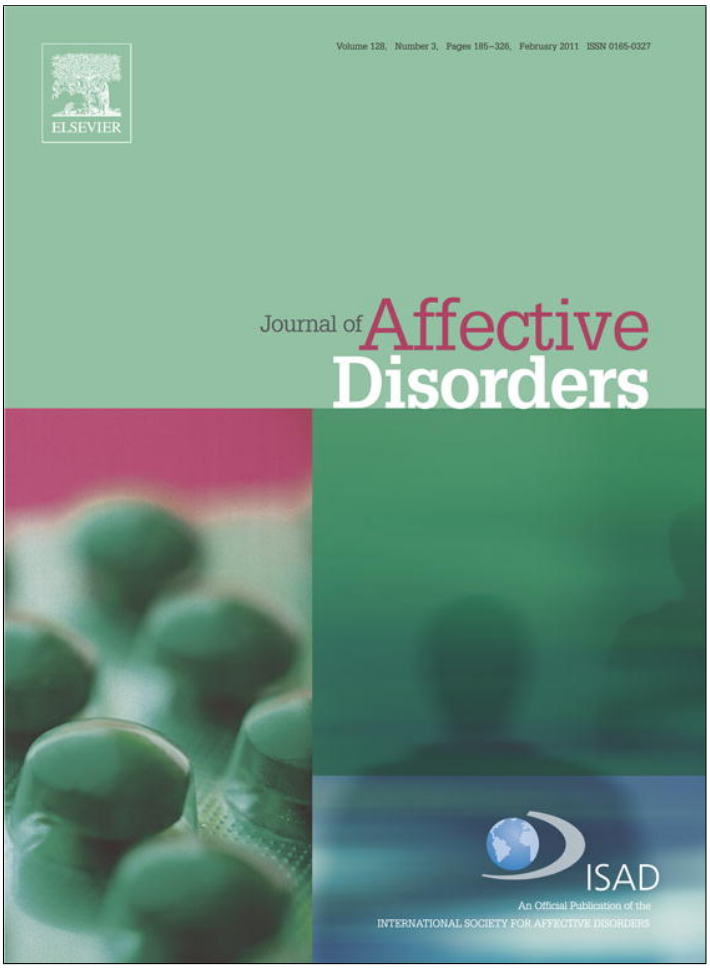
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.




This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Journal of Affective Disorders 128 (2011) 220–234

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research report

# A very short visual analog form of the Center for Epidemiologic Studies Depression Scale (CES-D) for the idiographic measurement of depression

Grégory Moullec [a,b,1], Christophe Maïano [c,*,1], Alexandre J.S. Morin [d,1], Johana Monthuy-Blanc [a], Lisa Rosello [e], Grégory Ninot [a]

[a] University of Montpellier I, EA 4206 "Addictive, Performance and Health Behaviors", Montpellier, France
[b] Hôpital du Sacré-Coeur de Montréal, Research Center, Montreal, Quebec, Canada
[c] UMR 6233 "Institute of Movement Sciences, Etienne-Jules Marey", CNRS-University of Aix-Marseille II, Marseille, France
[d] University of Sherbrooke, Department of Psychology, Sherbrooke, Québec, Canada
[e] Montfavet Psychiatric Hospital Center, Avignon, France

ARTICLE INFO

*Article history:*
Received 7 April 2010
Received in revised form 2 June 2010
Accepted 3 June 2010
Available online 6 July 2010

*Keywords:*
Major depressive episode
Affect
Clinical
Community
Gender
Invariance

ABSTRACT

*Background:* The experience sampling method, also referred to as ecological momentary assessment (ESM-EMA) has recently gained popularity in the study of depression. However, no psychometrically sound multidimensional depression questionnaires specifically designed for the ESM-EMA context are currently available.
*Aims:* The main objective of the present study was to develop and validate a very short visual analog scale of the Center for Epidemiologic Studies Depression Scales (CES-D-VAS-VS) specifically designed for the ESM-EMA context. To this end, the full French version of the CES-D was adapted for the ESM-EMA context. From this full-length adapted version a very short version was then extracted from this longer instrument and validated.
*Study design:* A sample comprising 163 patients with a major depressive episode (MDE) and 306 participants without mental disorders was involved in this study.
*Results:* The obtained results provided support for the factor validity, strong measurement invariance (invariance of the loadings and intercepts of the measurement model) across sex and clinical status groups, reliability and convergent validity of the CES-D-VAS-VS. This instrument comprises 4 items measuring positive affect, depressive affect, somatic complaints and disturbed interpersonal relationships.
*Conclusion:* The present results provide preliminary evidence regarding the construct validity of the CES-D-VAS-VS among patients and community adults sample but also underline the need to rely on latent variables methods in the use of this instrument to account for the differential levels of measurement errors (uniquenessess) that were observed across groups.
© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Over the past decades, the use of idiographic data collection methods based on frequent repeated measurements (e.g., hourly, daily or weekly) of symptoms in natural or clinical settings has considerably increased in clinical psychology and psychiatric research (Ebner-Priemer and Trull, 2009a; Moskowitz et al., 2009; Shiffman et al., 2008; Thiele et al., 2002). These methods, alternatively referred to as *experience sampling methods* (ESM; Larson and Csikszentmihalyi, 1983) or *ecological momentary assessment* (EMA; Stone and Shiffman, 1994), are designed to track the momentary fluctuations of participants' feelings, behaviors and experiences in real time settings (Shiffman et al., 2008). In ESM-EMA methods, the measures are taken either through an event-based monitoring process (i.e., each time a

* Corresponding author. UMR 6233 "Institute of Movement Sciences, Etienne-Jules Marey", 163 Avenue de Luminy, CP 910, 13288 Marseille Cedex 9, France. Tel.: +33 491 759 653; fax: +33 491 170 415.
*E-mail address:* christophe.maiano@univmed.fr (C. Maïano).
[1] Since all 3 contributed equally to the preparation of this paper, the order of appearance of the first, second and third authors (G.M., C.M. and A.J.M.) was determined at random: they should be considered first authors.

0165-0327/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jad.2010.06.006

Author's personal copy



predefined target event occurs), or through a time-based process (i.e., according to predetermined or random time schedules). Participants' answers can be recorded on paper-and-pencil questionnaires presented in a diary format or on hand-held computer-assisted devices (i.e., personal digital assistant, smart phone, notebook, etc.).

In patients suffering from major depressive disorders, ESM-EMA methods have been used mostly to gain a better understanding of the (i) relationships between various risk factors and short term fluctuations in the course of depressive symptoms; and (ii) the effectiveness of therapeutic interventions on short term fluctuations in the course of depressive symptoms (e.g., Barge-Schaapveld and Nicolson, 2002; Barge-Schaapveld et al., 1995; Golier et al., 2001; Gordijn et al., 1994; Myin-Germeys et al., 2003; Peeters et al., 2006, 2003; Putnam and McSweeney, 2008). However, these studies essentially relied on the measurement of mood states (i.e., positive and negative affects) or on depression questionnaires that were not originally designed to be used within an ESM-EMA context. Indeed, traditional depression questionnaires: (i) are often too long for idiographic protocols in which they have to be completed repeatedly; (ii) generally assess the presence of stable depressive symptoms that have been present for some time (e.g., over the last two weeks); and (iii) tend to rely on ordered categorical answer scales (e.g., 4-point Likert scales) that are inadequate for capturing momentary fluctuations.

Conversely ESM-EMA methods require (i) short instruments that may be used repeatedly and quickly, yet validly, to assess depressive symptoms fluctuations without representing too much of a burden for the participants in order to avoid risks of non-compliance or of automatic responses; (ii) items that are formulated to reflect present momentary states; and (iii) answer scales allowing for the accurate evaluation of those momentary states (Ebner-Priemer and Trull, 2009a; Shiffman et al., 2008; Thiele et al., 2002). Although it could be argued that various very short (i.e., one to three items) depression instruments are presently available, they remain unsuitable for the ESM-EMA methods (for a review see Mitchell and Coyne, 2007). Indeed, all of these instruments have been developed for screening purposes and essentially assess depressed mood (e.g., “Are you depressed?”, “Do you often feel sad or depressed?”) or loss of interest/pleasure (e.g., “Have you lost interest”, “Have you lost pleasure?”) on dichotomous answer scales (i.e., yes/no, presence/absence of symptom) that are clearly problematic for the precise assessment of momentary fluctuations. In addition, notwithstanding the debate regarding the continuous or categorical nature of major depression, depressive symptoms clearly occur on a severity continuum rather than as a present-absent dichotomy (e.g., Flett et al., 1997; Hankin et al., 2005; Judd, 1997; Pickles and Angold, 2003). The lack of available standardized and psychometrically valid multidimensional depression questionnaires specifically designed for this methodology represent one of the biggest obstacles to ESM-EMA studies of depressive symptoms and the most serious limitation of previous studies based on this methodology. In addition, this lack of short multidimensional instruments probably explains why so many of the preceding studies solely relied on mood based inventories, thus neglecting the full multidimensionality of depressive disorders.

The purpose of the present study is to develop and validate a very short (4-item; one item per dimension of the original instrument) visual analog version of the French Center for Epidemiologic Studies Depression Scale (Radloff, 1977) specifically designed for ESM-EMA. The CES-D was chosen because it (i) evaluates the main symptoms of depression: positive affect (e.g., hopeful, happy, etc.), depressed affect (e.g., blues, sad, etc.), somatic complaints (e.g., bothered, appetite, etc.), and disturbed interpersonal relationships (e.g., unfriendly, disliked, etc.); and (ii) has been widely adapted and validated cross-culturally in China (Cheung and Bagley, 1998), France (Führer and Rouillon, 1989), Germany (Hautzinger, 1988), Greece (Fountoulakis et al., 2001), Italy (Fava, 1983), The Netherlands (Beekman et al., 1994), Portugal (Gonçalves and Fagulha, 2004), Russia (Dershem et al., 1996), Spain (Vazquez et al., 2007), etc. Finally, the decision was made to rely on a Visual Analog Scale (VAS), requiring respondents to mark their answers on a continuous line instead of a traditional ordinal rating scale, because VASs (e.g., McCormack et al., 1988): (i) allow for a sufficient level of sensitivity to daily or hourly fluctuations; (ii) do not impose artificial categories on the underlying continuous response; (iii) are less vulnerable to learning effects and consistency biases since respondents cannot remember their previous answers and repeat them to appear coherent; (iv) are quick and easy to administer, requiring little effort from the less motivated respondents; (v) are suitable for frequent and repeated use; and (vi) are characterized by a discriminating capacity superior to other types of answering format.

The development of this short form of the CES-D follows recent recommendations for the development of short form tests (for more details see Marsh et al., 2005; Smith et al., 2000) indicating that: (i) the short form should retains the content coverage of each factor; (ii) the short form should retains the same factor structure as the original instrument; (iii) the factor structure of the short form should provide goodness of fit indices that meet acceptable standards; (iv) the short form should provide satisfactory levels of reliability; (v) the short form should be strongly correlated to the original version of the instrument and to additional instruments measuring similar and related characteristics (in this study, we retained known correlates of depression: anxiety, hopelessness, and self-esteem); and (vi) the criterion-related validity should be similar between the short form and the original instrument.

In addition to these, two additional verifications were conducted. First, the measurement equivalence (i.e., invariance) of the resulting instrument across samples of clinically depressive and community subjects will be verified. The fact that none of the preceding studies conducted on the different versions of the CES-D that we were able to locate (Cheung and Bagley, 1998; Fountoulakis et al., 2001; Gonçalves and Fagulha, 2004; Knight et al., 1997; Radloff, 1977; Vazquez et al., 2007) provided evidence of this form of invariance is alarming given the fact that the CES-D is specifically designed to identify *clinical* depression in epidemiological *community* samples. However, to do so requires the preliminary verification that the CES-D does measure the same construct, in the same manner, notwithstanding the clinical/community status of the evaluated individuals (Vandenberg and Lance,

Author's personal copy



2000). In other words, measurement invariance tests allow to verify if the higher scores on the instrument – that should be observed in clinical samples of depressed individuals – are really due to higher levels on the construct of interest (i.e., depression) or to the instrument measuring a different construct, or measuring it differently in a clinical population (Meredith and Teresi, 2006). Such measurement bias could be present when (i) the items measure the construct with more or less error in the different subgroups (i.e., uniquenesses/errors non-invariance), (ii) the items are scored systematically higher or lower in the various subgroups irrespective of participant's level on the latent construct of interest (i.e., intercepts non-invariance), or (iii) the items are differently related to the construct of interest in the various subgroups (i.e., factor loadings non-invariance).

Second, the measurement invariance of the resulting instrument across genders will also be verified. Indeed, the observation that women present a rate of depression twice higher than men has repeatedly been called one of the best-known facts of psychiatric epidemiology (Angold and Worthman, 1993). One possible explanation for gender-based differences in depressive symptoms is that they are not "real" and are rather the result of one or more artifacts (Leach et al., 2008). Nevertheless, these artifact explanations were not supported in the context of empirical studies (Bebbington, 1996; Hankin and Abramson, 1999; Hyde et al., 2008; Piccinelli and Wilkinson, 2000; Sprock and Yoder, 1997; Wolk and Weissman, 1995). The hypothesis that the items commonly used in the CES-D could be gender-biased also recently received increased attention from epidemiologists and psychologists. Indeed, since 1993, five studies investigated potential gender biases in the CES-D (Cole et al., 2000; Gelin and Zumbo, 2003; Grayson et al., 2000; Stommel et al., 1993; Yang and Jones, 2007). Although they relied on different methodologies, these studies suggest that, given similar levels of depression, women were likely to score higher (i.e., intercept non invariance) than men on some items [item 17 — "I had crying spells" (Cole et al., 2000; Gelin and Zumbo, 2003; Grayson et al., 2000; Stommel et al., 1993; Yang and Jones, 2007); item 10 — "I felt fearful" (Grayson et al., 2000; Yang and Jones, 2007); item 11 — "My sleep was restless" (Grayson et al., 2000; Yang and Jones, 2007)], while men were more likely to score higher on item 13 ["I talked less than usual"(Stommel et al., 1993)]. However, only one of those studies relied on a CFA methodology (Stommel et al., 1993). It is interesting to note that this study found no evidence of non-invariance in additional model parameters (loadings and uniquenesses).

## 2. Method

### 2.1. Participants and procedures

A total of 469 participants were involved in this study (65.7% women; $M_{age} = 40.7$, $SD_{age} = 16.2$, $range_{age} = 18–89$). This sample comprised a first sub sample of 163 patients (77.3% women; $M_{age} = 50.6$, $SD_{age} = 15.1$, $range_{age} = 19–89$) suffering from a Major Depressive Episode (MDE) according to the DSM-IV (American Psychiatric Association, 1994) and ICD-10 (World Health Organization, 2000) criteria. The second sub sample consisted of 306 community adults without any mental disorder or MDE (59.5% women, $M_{age} = 35.4$, $SD_{age} = 14.3$; $range_{age} = 18–82$). The clinical sample comprised participants which were on the average significantly older than those from the community sample [$t(467) = 10.78$, $p < .001$, $d = 1.05$].

The first subsample was recruited from two clinics (i.e., Costière and Saint Luc) and one inpatient unit in a Psychiatric Hospital (i.e., Montfavet) from southern France. Clinical Diagnosis of the patients was reached with the fifth version of the French Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). Only patients with a diagnostic of MDE (single or recurrent) on the MINI were included. Of the eligible patients, those with alcohol addiction and/or psychotic disorders according to DSM-IV and ICD-10 criteria were excluded from the study.

The second subsample comprised volunteer adults from southern France (i.e., Avignon, Nice, Marseille and Montpellier) that were recruited in the context of (i) university classes and (ii) student families. A brief interview with the volunteers was first conducted by a member from the research team and followed by the administration of sections of the MINI. This procedure was used to confirm that all participants were physically healthy, did not suffer from a MDE, and did not suffer from any other mental disorder. The volunteers who failed to meet these criteria were excluded from the study. Finally, all participants gave written informed consent, and the study protocol was approved by the ethical committee of the University of Montpellier I. All questionnaires used in the present study were administered by members of the research team in one-on-one sessions. In the visual analog version of the CES-D (CES-D-VAS), the participants were asked to indicate the extent to which the items characterized their present state by placing a vertical line intersecting each VAS ranging from "not at all" to "absolutely".

### 2.2. Measures

#### 2.2.1. Clinical diagnostic

The presence of a MDE diagnosis was assessed with the fifth version of the French MINI (Lecrubier et al., 1997; Sheehan et al., 1997, 1998). This instrument is a short, structured diagnostic interview that can be used as a tool to diagnose 16 Axis I psychiatric disorders according to DSM-IV and ICD-10 criteria. The MINI has 16 separate modules (e.g., MDE, anxiety, suicidal ideation, conduct disorders) each involving standardized, structured, close-ended questions. Interviewers read these questions verbatim to the interviewees. Psychiatric diagnosis and history in each specific module were made according to the number of affirmative replies to each question.

#### 2.2.2. Depressive symptoms

Three instruments were used to assess participant's levels of depressive symptoms: the original French version of the CES-D (Führer and Rouillon, 1989; Radloff, 1977), the VAS version of the French CES-D developed for this study, and the French version of the 13-item Beck Depression Inventory (BDI-13; Beck and Beck, 1972; Bourque and Beaudette, 1982; Collet and Cottraux, 1986).

The French version of the CES-D includes 20 items assessing the main symptoms of depression. These items are grouped into four distinct subscales, which converged on a single higher-order

Author's personal copy



factor of depression: (i) depressed affect (DA; 7 items), (ii) positive affect (PA; 4 items), (iii) somatic complaints (SC; 7 items), and (iv) disturbed interpersonal relationships (IP; 2 items). The answers to each item are given on a four-point scale on which the participant is asked to indicate the frequency with which he or she experienced the corresponding symptom during the past week [0 = rarely or none of the time (less than days); 1 = some or little of the time (1 to 2 days); 2 = occasionally or a moderate amount of the time (3 to 4 days); 3 = most or all of the time (5 to 7 days)]. Four of these items are reversed-scored to break possible answering tendencies. In this study the internal consistency coefficients were all in the acceptable range (CES-D: $\alpha = .92$ for DA, .78 for PA, .86 for SC, .70 for IP, and .94 for full scale) and comparable to those found in the original validation study of the French version of this instrument.

The original CES-D items were used to develop a preliminary VAS version of the full 20-items CES-D (CES-D-VAS). The content analysis of these items by four depression experts revealed that, in order to measure more adequately momentary fluctuations in current levels (i.e., at the moment) of depressive symptoms, (i) 18 items had to be slightly re-worded or simplified (i.e., items 1, 2, 4–16, 18–20); and (iii) 2 items (i.e., 3, 17) needed to be changed significantly. The final VAS items are reported in Table 1. The VAS scale that is used comprises a continuous 10 centimeters horizontal line defined by one verbal anchor at each extremity of the scale: « not at all» (0 cm) and «absolutely» (10 cm). Participants mark their answers by tracing a perpendicular mark on this scale.

The French version of the BDI-13 comprise 13 items rated on a behaviorally anchored scale ranging from 0 (absence of symptoms) to 3 (most severe symptoms) assessing the severity of depressive symptoms experienced during the past week including today. The French version of the BDI-13 has previously been shown satisfactory levels of internal consistency in a community sample ($\alpha = .90$), moderate test-retest correlations over a four months period ($r = .62$), and modest convergent validity with the Hamilton depression rating scale (Bourque and Beaudette, 1982; Collet and Cottraux, 1986). In this study the internal consistency of the BDI is satisfactory ($\alpha = .93$).

### 2.2.3. Anxiety

The French version of the Beck Anxiety Inventory (BAI; Beck et al., 1988; Freeston et al., 1994) was used to assess participants' levels of anxiety. The 21 items are answered on a 4-point Likert scale ranging from 0 (not at all) to 3 (severely, I could barely stand it) on which the respondents indicate the degree to which they have been bothered by each symptom during the past week including today. The French version of the BAI has previously been shown to present satisfactory levels of internal consistency (ranging from $\alpha = .84$ to .93 across four community samples), moderate test-retest

**Table 1**
Items of CES-D-VAS.

| Number | Original items | VAS items | Scale |
|---|---|---|---|
| 1 | J'ai été contrarié(e) par des choses qui d'habitude ne me dérangent pas. (I was bothered by things that usually don't bother me.) | Je suis dérangé(e) par des choses qui habituellement ne me dérangent pas. (*I'm bothered by things that usually don't bother me.*) | SC |
| 2 | Je n'ai pas eu envie de manger, j'ai manqué d'appétit. (I did not feel like eating; my appetite was poor.) | Je n'ai pas envie de manger, je n'ai pas d'appétit. (*I do not feel like eating; my appetite is poor.*) | SC |
| 3 | J'ai eu l'impression que je ne pouvais pas sortir du cafard, même avec l'aide de ma famille ou de mes ami(e)s. (I felt that I could not shake off the blues even with help from my family or friends.) | Je ne peux me défaire de mes idées noires même avec l'aide de ma famille ou de mes amis. (*I feel that I cannot shake off the blues even with help from my family or friends.*) | DA |
| 4 | J'ai eu le sentiment d'être aussi bien que les autres. *(I felt that I was just as good as other people.)* | Je me sens aussi bon que les autres. (*I feel that I am just as good as others.*) | PA* |
| 5 | J'ai eu du mal à me concentrer sur ce que je faisais. *(I had trouble keeping my mind on what I was doing.)* | J'ai de la difficulté à me concentrer sur ce que je fais. (*I have trouble keeping my mind on what I am doing.*) | SC |
| 6 | Je me suis senti(e) déprimée. *(I felt depressed.)* | Je me sens déprimé. (*I feel depressed.*) | DA |
| 7 | J'ai eu l'impression que toute action me demandait un effort. *(I felt that everything I did was an effort.)* | Tout ce que je fais me demande un effort. (*I feel that everything I do is an effort.*) | SC |
| 8 | J'ai été confiant(e) en l'avenir. *(I felt hopeful about the future.)* | Je me sens plein d'espoir en l'avenir. (*I feel hopeful about the future.*) | PA* |
| 9 | J'ai pensé que ma vie était un échec. *(I thought my life had been a failure.)* | Je sens que ma vie est un échec. (*I feel that my life is a failure.*) | DA |
| 10 | Je me suis senti(e) craintif(ve). *(I felt fearful.)* | Je me sens craintif. (*I feel fearful.*) | DA |
| 11 | Mon sommeil n'a pas été bon. *(My sleep was restless.)* | J'ai le sommeil agité. (*My sleep is restless.*) | SC |
| 12 | J'ai été heureux(se). *(I was happy.)* | Je suis heureux(se). (*I am happy.*) | PA†* |
| 13 | J'ai parlé moins que d'habitude. *(I talked less than usual.)* | Je parle moins que d'habitude. (*I talk less than usual.*) | SC† |
| 14 | Je me suis senti(e) seul(e). *(I felt lonely.)* | Je me sens seul(e). (*I feel lonely.*) | DA |
| 15 | Les autres ont été hostiles envers moi. *(People were unfriendly.)* | Les gens ne sont pas amicaux. (*People are unfriendly.*) | IP |
| 16 | J'ai profité de la vie. *(I enjoyed life.)* | Je profite de la vie. (*I enjoy life.*) | PA* |
| 17 | J'ai eu des crises de larmes. *(I had crying spells.)* | J'ai des crises de larmes ou me sens comme si j'allais éclater en sanglots. (*I have crying spells or feel like it.*) | DA† |
| 18 | Je me suis senti(e) triste. *(I felt sad.)* | Je me sens triste. (*I feel sad.*) | DA |
| 19 | J'ai eu l'impression que les gens ne m'aimaient pas. *(I felt that people disliked me.)* | Je sens que les gens ne m'aiment pas. (*I feel that people dislike me.*) | IP† |
| 20 | J'ai manqué d'entrain. *(I could not get "going".)* | Je ne parviens pas à aller de l'avant. (*I cannot get "going".*) | SC |

Note. CES-D: Center for Epidemiologic Studies-Depression scale; VAS: Visual analog Scale; DA: Depressed affect; PA: Positive affect; SC: Somatic complaints; IP: Disturbed interpersonal relationships; *: reversed score; †: Items that were retained in the very short form.

Author's personal copy

**Table 2**
Goodness of Fit Indices of CES-D-VAS and CES-D-VAS-VS Models[a].

| Stages | Model | N° | Description | $\chi^2$(B-S) | df | CFI | TLI | SRMR | RMSEA | RMSEA 90% CI | $\Delta\chi^2$(df) | ΔCFI | ΔTLI | ΔRMSEA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stage 1 | CFA | 1 | CES-D-VAS: Single first-order | 266.806* | 170 | .904 | .892 | .047 | .083 | .077–.090 | | | | |
| | | 2 | CES-D-VAS: Two single order (DA-SC; PA-IP) | 264.276* | 169 | .918 | .904 | .050 | .077 | .071–.083 | | | | |
| | | 3 | CES-D-VAS: Three first order (PA-DA, SC, IP) | 261.558* | 167 | .925 | .914 | .042 | .074 | .068–.081 | | | | |
| | | 4 | CES-D-VAS: Three first-order (DA-SC, PA, IP) | 260.791* | 167 | .944 | .937 | .041 | .064 | .057-.070 | | | | |
| | | 5 | CES-D-VAS: Four first-order | 255.079* | 164 | .951 | .943 | .041 | .060 | .054–.067 | | | | |
| | | 6 | CES-D-VAS: Single second-order and four first-order | 258.466* | 166 | .951 | .943 | .040 | .060 | .054–.067 | | | | |
| Stage 2 | CFA, gender-invariance tests - first order | 6A | 1-Configural invariance | 539.656* | 328 | .924 | .912 | .066 | .053 | .048–.058 | | | | |
| | | | 2-λ invariant | 559.114* | 344 | .923 | .915 | .073 | .052 | .047–.057 | 19.46(16) | −.001 | +.003 | −.001 |
| | | | 3-λ, τs invariant | 575.218* | 360 | .918 | .914 | .071 | .053 | .048–.057 | 16.10(16) | −.005 | −.001 | +.001 |
| | | | 4-λ, τs, δs invariant | 626.348* | 380 | .882 | .882 | .112 | .062 | .057–.066 | 51.13(20)* | −.036 | −.032 | +.009 |
| | | | 4′-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free) invariant | 587.540* | 366 | .917 | .914 | .075 | .053 | .048–.057 | 12.32(6) | −.001 | .000 | .000 |
| | | | 5-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), ξ/φ invariant | 604.009* | 376 | .910 | .909 | .090 | .054 | .049–.059 | 16.47(10) | −.007 | −.005 | +.001 |
| | | | 6-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), ξ/φ, $\eta^{1rd}$s invariant | 607.631* | 380 | .905 | .905 | .097 | .055 | .051–.060 | 3.62(4) | −.005 | −.004 | +.001 |
| | CFA, gender-invariance tests—second order | 6B | 1-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free) invariant | 594.205* | 370 | .917 | .915 | .076 | .052 | .048–.057 | | | | |
| | | | 2-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), γs invariant | 597.797* | 373 | .916 | .915 | .075 | .052 | .048–.057 | 3.60(3) | −.001 | .000 | .000 |
| | | | 3-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), γs, $\eta^{1rd}$ invariant | 601,064* | 376 | .915 | .915 | .075 | .052 | .048–.057 | 3.27(3) | −.001 | .000 | .000 |
| | | | 4-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), γs, $\eta^{1rd}$, ζs invariant | 609.546* | 380 | .911 | .911 | .084 | .054 | .049–.058 | 8.48(4) | −.004 | −.004 | +.002 |
| | | | 5-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), γs, $\eta^{1rd}$, ζs, $\xi^{2rd}$ invariant | 612.581* | 382 | .877 | .878 | .112 | .063 | .058–.067 | 3.04(2) | −.034 | −.033 | +.009 |
| | | | 5′-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), γs, $\eta^{1rd}$, ζs, $\xi^{2rd}$ (free) invariant | 609.546* | 380 | .911 | .911 | .084 | .054 | .049–.058 | 0.00(0) | .000 | .000 | .000 |
| | | | 6-λ, τs, δs (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19, 20 free), γs, $\eta^{1rd}$, ζs, $\xi^{2rd}$ (free), $\eta^{2rd}$ invariant | 610.208* | 381 | .905 | .906 | .080 | .055 | .051–.060 | 0.66(1) | −.006 | −.005 | +.001 |
| | CFA, clinical-invariance tests—first order | 6C | 1-Configural invariance | 477.882* | 328 | .889 | .871 | .069 | .050 | .045–.055 | | | | |
| | | | 2-λ invariant | 497.323* | 344 | .882 | .870 | .080 | .050 | .045–.055 | 19.44(16) | −.007 | −.001 | .000 |
| | | | 3-λ, τs invariant | 520.510* | 360 | .869 | .862 | .075 | .051 | .047–.056 | 23.19(16) | −.013 | −.008 | +.001 |
| | | | 3′-λ, τs (8, 11 free) invariant | 518.570* | 358 | .879 | .872 | .076 | .050 | .045–.055 | 21.25(14) | −.003 | +.002 | .000 |
| | | | 4-λ, τs (8, 11 free), δs invariant | 559.118* | 378 | .637 | .635 | .114 | .084 | .080–.088 | 40.55(20)* | −.242 | −.237 | +.034 |
| | | | 4′-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free) invariant | 522.145* | 360 | .875 | .868 | .077 | .050 | .046–.055 | 3.57(2) | −.004 | −.004 | .000 |
| | | | 5-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), ξ/φ invariant | 535.623* | 370 | .855 | .851 | .113 | .054 | .049–.058 | 13.45(10) | −.020 | −.017 | +.004 |
| | | | 5′-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), ξ/φ (ξ IP, PA free) invariant | 532.782* | 368 | .866 | .861 | .114 | .052 | .047–.056 | 10.64(8) | −.009 | −.007 | +.002 |
| | | | 6-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), ξ/φ (ξ IP, PA free), $\eta^{1rd}$s invariant | 529.901* | 372 | .765 | .760 | .106 | .068 | .064–.072 | −2.88(6) | −.101 | −.101 | +.016 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CFA, clinical-invariance tests, second-order | 6D | 1-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free) invariant | 528.145* | 364 | .875 | .870 | .077 | .050 | .045–.055 | | | | |
| | | | 2-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs invariant | 532.119* | 367 | .875 | .871 | .079 | .050 | .045–.055 | 3.97(3) | .000 | +.001 | .000 |
| | | | 3-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs, $\eta^{1rd}$ invariant | 536.824* | 370 | .875 | .872 | .080 | .050 | .045–.054 | 4.71(3) | .000 | +.001 | .000 |
| | | | 4-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs, $\eta^{1rd}$, ζs invariant | 542.841* | 374 | .863 | .861 | .086 | .052 | .047–.056 | 6.02(4) | −.012 | −.011 | +.002 |
| | | | 4′-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs, $\eta^{1rd}$, ζs (IP, PA free) invariant | 539.715* | 372 | .874 | .872 | .081 | .050 | .045–.054 | 2.89(2) | −.001 | .000 | .000 |
| | | | 5-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs, $\eta^{1rd}$, ζs (IP, PA free), $\xi^{2rd}$s invariant | 542.525* | 374 | .846 | .844 | .151 | .055 | .050–.059 | 2.81(2) | −.028 | −.028 | −.005 |
| | | | 5′-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs, $\eta^{1rd}$, ζs (IP, PA free), $\xi^{2rd}$ (free) invariant | 539.715* | 372 | .874 | .872 | .081 | .050 | .045–.054 | 0.00(0) | .000 | .000 | .000 |
| | | | 6-λ, τs (8, 11 free), δs (1–4, 6–15, 17–20 free), γs, $\eta^{1rd}$, ζs (IP, PA free), $\xi^{2rd}$ (free), $\eta^{2rd}$ invariant | 537.149* | 373 | .811 | .807 | .320 | .061 | .056–.065 | −2.57(1) | −.063 | −.065 | −.011 |
| Stage 3 | CFA | 7 | CES-D-VAS-VS: Single 4-item factor | 2.770 | 2 | .990 | .971 | .025 | .068 | .011–.131 | | | | |
| | CFA, gender-invariance tests | 7A | 1- Configural invariance | 6.085 | 4 | .984 | .952 | .021 | .062 | .019–.106 | | | | |
| | | | 2-λ invariant | 10.846 | 7 | .977 | .961 | .042 | .055 | .021–.089 | 4.76(3) | −.007 | +.009 | −.007 |
| | | | 3-λ, τs invariant | 13.843 | 10 | .969 | .963 | .042 | .054 | .026–.083 | 3.00(3) | −.008 | +.002 | −.001 |
| | | | 4-λ, τs, δs invariant | 24.312* | 14 | .771 | .804 | .195 | .124 | .103–.145 | 10.47(4)* | −.198 | −.159 | +.070 |
| | | | 4′-λ, τs, δs (12, 17, 19 free) invariant | 15.894 | 11 | .970 | .967 | .040 | .051 | .023–.078 | 2.05(1) | +.001 | +.004 | −.003 |
| | | | 5-λ, τs, δs (12, 17, 19 free), ξs invariant | 18.153 | 12 | .953 | .953 | .089 | .060 | .036–.086 | 2.26(1) | −.017 | −.014 | +.009 |
| | | | 5′-λ, τs, δs (12, 17, 19 free), ξs (free) invariant | 15.894 | 11 | .970 | .967 | .040 | .051 | .023–.078 | 0.00(0) | 0.00 | 0.00 | 0.00 |
| | | | 6-λ, τs, δs (12, 17, 19 free), ξs (free), η invariant | 18.836* | 13 | .862 | .872 | .117 | .100 | .078–.123 | 2.94(2) | −.108 | −.095 | +.049 |
| | CFA, clinical-invariance tests | 7B | 1-Configural invariance | 5.037 | 4 | .998 | .993 | .044 | .014 | .000–.073 | | | | |
| | | | 2-λ invariant | 8.959 | 7 | 1.00 | 1.00 | .064 | .000 | .000–.056 | 3.92(3) | +.002 | +.007 | −.014 |
| | | | 3-λ, τs invariant | 14.621 | 10 | .993 | .992 | .068 | .015 | .000–.054 | 5.66(3) | −.007 | −.008 | +.015 |
| | | | 4-λ, τs, δs invariant | 21.150 | 14 | Model failed to converge on a proper solution | | | | | | | | |
| | | | 4′-λ, τs, δs (12, 13, 17, 19 free) invariant | 14.621 | 10 | .993 | .992 | .068 | .015 | .000–.054 | 5.66(3) | 0.00 | 0.00 | 0.00 |
| | | | 5-λ, τs, δs (12, 13, 17, 19 free), ξs invariant | 15.729 | 11 | .980 | .978 | .097 | .026 | .000–.059 | 1.11(1) | −.013 | −.014 | +.011 |
| | | | 5′-λ, τs, δs (12, 13, 17, 19 free), ξs (free) invariant | 14.621 | 10 | .993 | .992 | .068 | .015 | .000–.054 | 5.66(3) | 0.00 | 0.00 | 0.00 |
| | | | 6-λ, τs, δs (12, 13, 17, 19 free), ξs (free), η invariant | 14.935* | 11 | Model failed to converge on a proper solution | | | | | | | | |

Note. CFA: Confirmatory factor analytic model; CES-D: Center for Epidemiologic Studies-Depression scale; VAS: Visual analog scale; VS: Very Short form; $\chi^2$ (B–S): Bollen–Stine chi-square; df: Degrees of freedom; CFI: Comparative fit index; TLI: Tucker–Lewis index; SRMR: Standardized root mean square residual; RMSEA: Root mean square error of approximation; RMSEA 90% CI = 90% Confidence interval for the RMSEA point estimate; λ: Factor loading; τ: Intercept; δ: Uniquenesses; ξ: Factor variance; $\xi^{2rd}$: Second-order factor variance; φ: Factor covariance; γ: Structural relations among latent constructs; ζ: Factor error terms; η: factor mean; $\eta^{1rd}$: First-order factor means; $\eta^{2rd}$: Second-order factor means;Δ$\chi^2$: Change in goodness-of-fit $\chi^2$ relative to the preceding model; Δdf: Change in degrees of freedom relative to the preceding model; ΔCFI: Change in comparative fit index relative to the preceding model; ΔTLI: Change in Tucker–Lewis index relative to the preceding model; ΔRMSEA: Change in root mean square error of approximation relative to the preceding model; [a] Bootstrapped goodness of fit indexes are reported in this table because of the significant multivariate non-normality within these data.

* $p < .05$.



Author's personal copy

Author's personal copy



correlations over a four weeks period (r = .63), and satisfactory levels of convergent validity with measures of depression, fear of negative evaluation, irrational beliefs, obsessive-compulsive complaints and assertive behaviors (Freeston et al., 1994). In this study the internal consistency of the BAI is satisfactory (α = .93).

*2.2.4. Hopelessness*

The French version of the Beck Hopelessness Scale (BHS; Beck et al., 1974; Bouvard et al., 1992) was used to measure negative attitudes about the future experienced by the respondents over the past week. This self-report instrument consists of 20 true-false statements which are scored 0 or 1. The French version of the BHS has previously been shown to present satisfactory levels of internal consistency in clinically depressed (α = .89) and community (α = .79) samples, elevated test–retest correlations over a two weeks period (r = .81), and satisfactory levels of convergent validity with measures of dysfunctional attitudes, suicidal ideation and automatic thoughts (Bouvard et al., 1992). In this study the internal consistency of the BHS is satisfactory (α = .88).

*2.2.5. Self-esteem*

The French version of the Rosenberg Self-Esteem Inventory (RSEI; Rosenberg, 1965; Vallières and Vallérand, 1990) was used to assess overall feelings of self-worth or self-acceptance. The 10 items from this instrument are rated on a 4-point Likert scale ranging from *strongly agree* (4) to *strongly disagree* (1). The French version of the RSEI has previously been shown to present satisfactory levels of internal consistency (ranging from α = .70 to .90 across four community samples), elevated test–retest correlations over a three weeks period (r = .84), and satisfactory levels of convergent validity with measures of depression and life satisfaction (Vallières and Vallérand, 1990). In this study the internal consistency of the RSEI is in the acceptable range (α = .75).

*2.2.6. Analyses*

Confirmatory Factor Analysis (CFA) were performed using Bootstrapped Maximum Likelihood estimation with AMOS 7.0 program (Arbuckle, 2006), given the significant non-normality of depressive symptoms. Assessment of model fit and comparison between models were based on multiple indicators (Byrne, 2005; Hu and Bentler, 1999): the Chi-square statistic ($\chi^2$), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Standardized Root Mean square Residual (SRMR), the Root Mean Square Error of Approximation (RMSEA) and the 90% confidence interval of the RMSEA. Values greater than .90 for CFI and TLI are considered to be indicative of adequate model fit (Byrne, 2005; Hu and Bentler, 1999; Vandenberg and Lance, 2000), although values approaching .95 are preferable. Values smaller than .08 or .06 for the RMSEA and smaller than .10 and .08 for the SRMR support respectively acceptable and good model fit (Hu and Bentler, 1999; Vandenberg and Lance, 2000). Concerning the RMSEA 90% CI, values of less than .05 for the lower bound (left side) and less than .08 for the upper bounds (right side) or containing 0 for the lower bound and less .05 for the upper bounds (right side) indicate respectively acceptable and good model fit (MacCallum et al., 1996). Critical values for the tests of multi-group invariance in CFAs models were evaluated by the examination of several criterions, using the preceding model as the comparison model: $\chi^2$ difference tests and changes in CFI and RMSEA (Chen, 2007; Cheung and Rensvold, 2002; Vandenberg and Lance, 2000). A CFI difference of .01 or less and a RMSEA difference of .015 or less between a reference model and the following model indicate that the measurement invariance hypothesis should not be rejected. Vandenberg and Lance (2000) also indicate that CFI differences of .02 or more would be needed to clearly reject the measurement invariance hypothesis.

## 3. Results

### 3.1. Stage 1. Factor validity and reliability of the CES-D-VAS

Six *a priori* CFA models from the extant literature (for a review see Perreira et al., 2005; Shafer, 2006; Sheehan et al., 1995) were examined for the CES-D-VAS scores: (i) a single factor model (Model 1); (ii) a two factor model (Model 2: DA-SC; PA-IP); (iii) two different three factor models (Model 3: PA-DA, SC, IP; Model 4: DA-SC, PA, IP); (iv) a four correlated factor model (Model 5); (v) a four factor model with a single higher-order factor (Model 6). Model 1 a priori hypothesized that: (i) answers to the CES-D-VAS could be explained by a single factor of depression; (ii) each item would have a non-zero loading on the depression factor; and (iii) error terms would be uncorrelated. Models 2 to 6 a priori hypothesized: (i) answers to the CES-D-VAS could be explained by 2 to 4 first-order factors (previously defined); (ii) each item would have a non-zero loading on the CES-D-VAS factor it was designed to measure, and zero loadings on all other factors; (iii) the first-order factors would be correlated (Models 2 to 5) or load on a single higher-order factor of depression (Model 6); and (iv) error terms would be uncorrelated.

The goodness-of-fit statistics of the various CFA models tested for the CES-D-VAS are displayed in Table 2. They showed that Model 1 exhibited poor goodness of fit indices (i.e., TLI<.90; RMSEA>.08). Inversely, the other five CFA models (2 to 6), provided modest (Models 2–3: CFI and TLI>.90; SRMR<.06; RMSEA<.08) to acceptable levels of fit to the data (Models 4–6: CFI>.95; TLI<.90; SRMR<.05; RMSEA = .06). Among Models 4 to 6, the results support the superiority of models 5 and 6, which present an equivalent level of fit to the data. However, since the hierarchical model (Model 6) is more parsimonious (replacing six latent factors correlations by four second order factor loadings and thus freeing two degrees of freedom) as well as more convergent with the theoretical framework underlying the original CES-D, this model was retained for the development of the very short form of the CES-D-VAS, hereafter referred to as the CES-D-VAS-VS. The estimated standardized loadings from this model were all significant and substantial and are illustrated in Fig. 1. Finally, the reliability of the Model 6 was computed from the model's standardized parameters estimates, using the formula provided by McDonald (1999): $\omega = (\Sigma\lambda i)^2 / ([\Sigma\lambda i]^2 + \Sigma\delta ii)$ where $\lambda i$ are the factor loading and $\delta ii$ the error variances. Results revealed that the scales of this model present, for the pooled sample, modest to acceptable internal reliability coefficients ($\omega$) of .87 for DA, .79 for PA, .87 for SC, .67 for IP, and .95 for full scale.

Author's personal copy



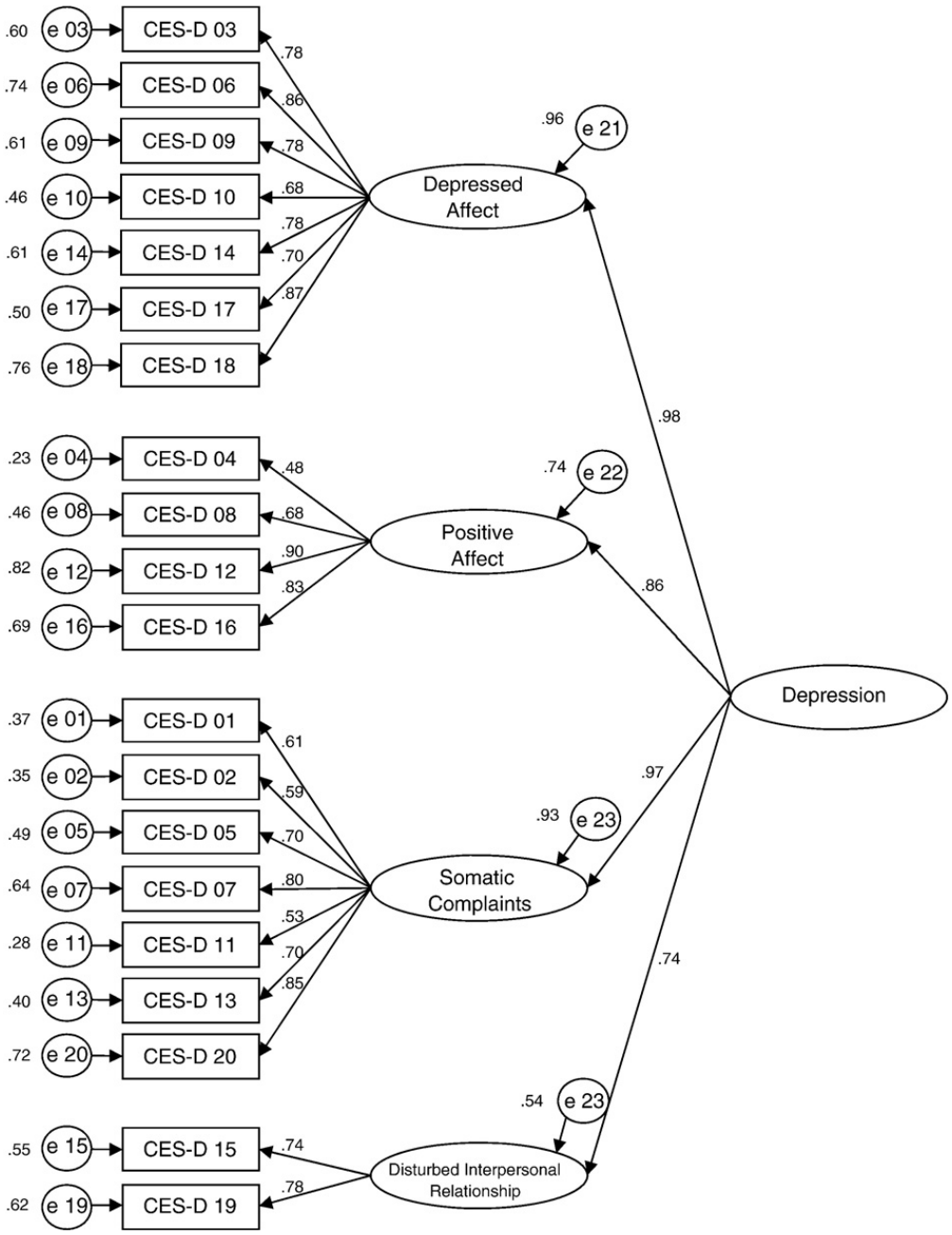



**Fig. 1.** Estimated standardized uniquenesses, disturbances and loadings for Model 6[a]. Note. [a]All estimated are significant ($p<0.001$).

### *3.2. Stage 2. Measurement and latent mean invariance of the CES-D-VAS*

In the second stage, the measurement invariance of the CES-D-VAS across gender (Model 6A and 6B) and clinical groups (Models 6C and 6D) was verified following a sequential strategy devised through a combination of Meredith (1993) recommendations for first-order factor models (Models 6A and 6C) and Cheung (2008) recommendations for higher-order factor models (Model 6B and 6D). The measurement invariance of the first-order factor model was thus estimated first, without a second-order latent construct (Cheung, 2008), in the following sequence: (i) configural invariance, (ii) loadings invariance (metric invariance), (iii) intercepts invariance (strong invariance), (iv) uniquenesses invariance (strict invariance), (v) variance/covariance invariance, and (vi) latent means invariance. Then, the invariance of the second-order structure was verified in the following sequence, with the baseline specified according to the conclusions of the steps (i) to (iv) of the preceding sequence: (i) baseline; (ii) second-order loadings invariance; (iii) second-order intercepts invariance; (iv) second-order disturbances invariance; (v) second-order variance invariance; and (vi) second-order latent mean invariance.

The results from the CFAs gender-based measurement and latent mean invariance tests for the first-order structure (Model 6A) revealed that the three first steps of invariance testing, including configural invariance and the invariance of the loadings and intercepts, resulted in significant bootstrap $\chi^2$, acceptable goodness of fit-indices and in equivalent fit indices (non significant $\Delta\chi^2$, $\Delta$CFIs, $\Delta$RMSEAs). The fourth level of measurement invariance added the equality

Author's personal copy



constraints on the items' uniquenesses. These results did not support the measurement invariance of the items' uniquenesses across gender (significant $\Delta\chi^2$ and $\Delta$CFI>.01), and equality constraints across gender had to be relaxed for the uniquenesses of approximately two thirds of the items (1–4, 6, 7, 9, 10, 12, 14, 16, 17, 19 and 20). These parameters were then freely estimated across gender (level 4′), while keeping the other constraints, and provided evidence of partial strict measurement invariance of the full model but of non invariance for a majority of the items from the DA (six out of seven items), PA (three out of four items) and SC (four of seven items) factors (as well as for item 19 of the IP factor). In other words, these results reveal that men and women rate these items with differing levels of measurement error. The next models verified the invariance of the variance/covariance matrix and of the latent means. These models resulted in significant bootstrap $\chi^2$, acceptable goodness of fit-indices and equivalent fit indices. These results suggest that when the constraints on the non-invariant uniquenesses are relaxed, women average levels and variability on the latent variables of interest do not differ from those of men.

The results from the subsequent CFAs in which the gender-based measurement and latent mean invariance of the second-order structure (Model 6B) was verified supported the complete measurement and latent mean invariance of the higher-order CFA model, with the exception of the equality constraints on the variance of the second-order factor that had to be relaxed across gender (level 5′), revealing slightly higher second-order variance in the women subsample. Once again, these results suggest that when the constraints on the non-invariant uniquenesses are relaxed, women average levels on the second order latent mean did not differ from men's levels.

The results from the CFA testing the measurement and latent mean invariance of the first-order factor structure across the clinical and non-clinical groups (Model 6C) revealed: (i) the complete invariance of the factor loadings; (ii) the partial invariance of the intercepts (i.e., equality constraints on item intercepts 8 and 11 had to be relaxed, revealing higher intercepts for the community group), uniquenesses (i.e., equality constraints had to be relaxed on the uniquenesses of all items except 5 and 16), and the factor variance/covariance matrix (i.e., equality constraints on the IP and PA factors variance had to be relaxed, revealing higher variance in the clinical group); and (iii) the non invariance of the first order factor means, with the clinical group (latent means freely estimated) presenting a significantly higher level than the community group (latent means fixed to zero) on the DA (latent mean = 4.06, $t = 16.25$, $p<.001$, $d = 1.58$), PA (latent mean = 2.37, $t = 9.45$, $p<.001$, $d = .92$), SC (latent mean = –3.06, $t = 12.61$, $p<.001$, $d = 1.23$) and IP (latent mean = 2.42, $t = 9.50$, $p<.001$, $d = .92$) factors.

Regarding the second-order factor structure (Model 6D), the results support: (i) the complete invariance of the factor loadings, intercepts and disturbances; (ii) the non invariance of the second-order factor variance (i.e., equality constraints on the variance of the depression global factor had to be relaxed, revealing higher variance for the clinical group); and (iii) the non invariance of the second-order factor means, with the clinical group (latent mean = 2.73, $t = 18.98$, $p<.001$, $d = 1.85$) presenting a significantly higher level of depression than the community group (latent mean fixed to zero).

### 3.3. Stage 3. Development and factor validity and reliability of the CES-D-VAS-VS

In the third stage, the data from Model 6 were used to develop a preliminary very short version of CES-D-VAS (CES-D-VAS-VS) by selecting one item by dimension. The items were retained on the basis of Marsh et al. (2005) and Smith et al. (2000) criteria. These items: (i) were those with the best item-total correlations and factor loadings; (ii) had minimal cross-loadings as evidenced by AMOS modifications indices; (iii) presented a low level of correlated uniquenesses; (iv) were very seldom left blank by the respondents; and (v) received positive subjective evaluations of their content from two experts from the depression field (i.e., two psychiatrists). On the basis of the aforementioned recommendations, the items 12, 13, 17 and 19 were selected.

Once the items of the CES-D-VAS-VS were selected, the data from the pooled sample were submitted to a single-factor CFA model (Model 7) in which it was hypothesized that: (i) the answers to the CES-D-VAS-VS could be explained by one single-depression factor; (ii) each item would have a non-zero loading on the single-depression factor; and (iii) measurement error terms, or uniquenesses, would be uncorrelated. It should be noted that a first order, one factor model, was used to validate the CES-D-VAS-VS to account for the fact that a single item was selected from each dimension of the original CES-D-VAS. The theoretical adequacy of this single factor model is supported by the fact that all items are purported to measure depression (a global scale score can be computed), an assumption that was confirmed by the selection of the second order factor model as the model providing the best fit to the data in the preceding analyses of the CES-D-VAS. Results from the first-order CFA performed on those 4 items of the CES-D-VAS-VS are provided in Table 2 and presented acceptable goodness of fit indices (CFI and TLI>.95; RMSEA<.07 and SRMR<.03). As illustrated in Fig. 2., all loadings in this CFA model were significant and substantial. This scale also presented a satisfactory level of reliability ($\omega = .80$).

### 3.4. Stage 4. Measurement and latent mean invariance of the CES-D-VAS-VS

Concerning the fourth stage, the measurement invariance tests across gender and clinical groups (Models 7A and 7B) were performed in the sequential order recommended by Meredith (1993): (i) configural invariance, (ii) factor loadings invariance (metric invariance), (iii) items intercepts

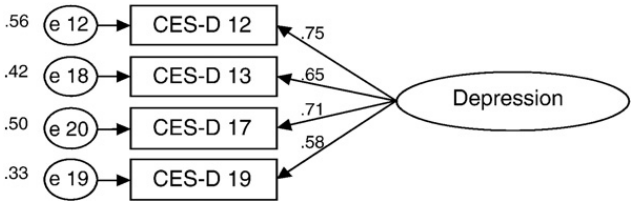



**Fig. 2.** Estimated standardized uniquenesses and loadings for Model 7[a]. Note. [a]All estimated are significant ($p<0.001$).

Author's personal copy



invariance (strong invariance), (iv) items uniquenesses invariance (strict invariance), (v) factor variance invariance, and (vi) factor mean invariance.

The results from the gender-based measurement and latent mean invariance tests are presented under Model 7A. The first three steps of invariance testing (i.e., configural invariance, invariance of the loadings and invariance of the intercepts) resulted in non-significant bootstrap $\chi^2$, acceptable goodness of fit-indices (i.e., CFI, TLI>.95; SRMR<.05; RMSEA<.08), and in equivalent goodness of fit indices (non significant $\Delta\chi^2$, ΔCFIs, ΔRMSEAs). The fourth level (i.e., equality constraints on items' uniquenesses) provided significant bootstrap $\chi^2$ and acceptable goodness of fit-indices. However, this model resulted in a significant $\Delta\chi^2$ and a ΔCFI and ΔRMSEA that did exceed .015. Further examination of these results revealed that the gender-group equality constraint need to be relaxed for the uniqueness of three items (12, 17 and 19), revealing a higher level of measurement error in women for items 12, 17 and 19. This result is perfectly consistent with the findings from the CES-D-VAS gender-based invariance tests in which most of the items uniquenesses already proved non-invariant across gender. These parameters were then freely estimated across gender groups while keeping the other constraints (level 4′) and provided evidence of partial strict measurement invariance. The next model (i.e., hypothesis 5: equality constraints on factor variance) provided non-significant bootstrap $\chi^2$, acceptable goodness of fit-indices, a non-significant $\Delta\chi^2$and ΔRMSEA that did not exceed .015. However the ΔCFI (−0.017) proved to be over the 0.01 and suggest the non invariance of the factor variance. When this equality constraint was relaxed (level 5′), the results showed that women presented a higher level of variability on this factor. The final model in which equality constraints were imposed on the latent factors means did not support the invariance of the latent means across gender (i.e., significant $\Delta\chi^2$ and ΔCFI and ΔRMSEA>.015), and revealed that men (latent mean = −1.77, $t = -6.78$, $p<.05$, $d = .65$) presented significantly lower levels of depression than women (latent mean fixed to zero).

The results from the CFAs testing the measurement and latent mean invariance across the clinical and community groups are presented under model 7B. These results confirm the invariance of the factor loadings and items' intercepts across both groups. However, the model in which invariance constraints were added on the items uniquenesses failed to converge on a proper solution, which suggests the inadequacy of this model and is perfectly consistent with the results from the CES-D-VAS models in which most of the items uniquenesses proved non-invariant across clinical and community groups. Indeed, a detailed examination of the results from the preceding models, as well as a series of ex post facto modifications in which invariance constraints were relaxed on one item uniqueness at a time revealed that the equality constraints on all of the items uniqueness had to be relaxed; revealing higher levels of measurement errors in the community group. The results from the last two steps revealed that the invariance of the depression factor variance and latent mean also had to be relaxed across both groups. Detailed examination of these results revealed differences that are consistent with what could be expected: participants from the clinical group (latent mean fixed to zero) presented more variance and a higher average level of depression than community participants (latent mean = −4.42, $t = -15.87$, $p<.001$, $d = 1.54$).

### 3.5. Stage 5: Convergent validity of the CES-D-VAS-VS and the CES-D

In the fifth stage, correlation analyses were performed in order to verify the convergent validity of the global and subscale scores of the original CES-D with the selected items and the full scale of the CES-D-VAS-VS. A Bonferroni correction was applied to minimize Type I error rate inflation (alpha error was thus set at 0.05/5 = 0.01 to account for the 5 comparisons being made) The results reported in Table 3 show that the items and full scale of the CES-D-VAS-VS were more highly correlated with their parent CES-D subscales than with other CES-D subscales (most of these correlations were positive and significant). These results thus confirm the convergent validity of the CES-D-VAS-VS, relative to the original CES-D.

### 3.6. Stage 6: Convergent validity of the CES-D-VAS-VS with criterion-related measures

Concerning the sixth stage, correlational analyses were performed in order to verify the convergent validity of the CES-D-VAS-VS with another measure of depression (BDI-13) and with measures of anxiety (BAI), hopelessness (BHS) and self-esteem (RSEI). A Bonferroni correction was applied to minimize Type I error rate inflation (alpha error was thus set at 0.05/4 = 0.01). The results are reported in Table 4 and revealed that the full scale and all items of the CES-D-VAS-VS were: (i) significantly and positively correlated with the BDI-

**Table 3**
Convergent and discriminant validity of the CES-D-VAS-VS with the original CES-D.

| Scales | PA-CES-D | DA-CES-D | IP-CES-D | SC-CES-D | Full-CES-D |
|---|---|---|---|---|---|
| Item 12-PA-VAS-VS | .66* | .73* | .40* | .69* | .76* |
| Item 17-DA-VAS-VS | .48* | .69* | .36* | .61* | .66* |
| Item 19-IP-VAS-VS | .36* | .50* | .62* | .53* | .55* |
| Item 13-SC-VAS-VS | .49* | .60* | .33* | .65* | .64* |
| Full CES-D-VAS-VS | .66* | .83* | .54* | .82* | .86* |

Note. CES-D: Center for Epidemiologic Studies-Depression scale; VAS: Visual analog scale; VS: Very Short form; Depressed affect; PA: Positive affect; SC: Somatic complaints; IP: Disturbed interpersonal relationships.
* $p<.001$.

Author's personal copy



**Table 4**
Convergent validity of the CES-D-VAS-VS with criterion measures.

| Scales | BDI-13 | RSEI | BAI | BHS |
|---|---|---|---|---|
| Item 12-PA-VAS-VS | .73* | −.62* | .58* | .63* |
| Item 17-DA-VAS-VS | .61* | −.51* | .57* | .44* |
| Item 19-IP-VAS-VS | .56* | −.50* | .48* | .45* |
| Item 13-SC-VAS-VS | .59* | −.52* | .50* | .43* |
| Full CES-D-VAS-VS | .81* | −.70* | .70* | .64* |

Note. CES-D: Center for Epidemiologic Studies-Depression scale; VAS: Visual analog scale; VS: Very Short form; DA: Depressed affect; PA: Positive affect; SC: Somatic complaints; IP: Disturbed interpersonal relationships; BDI-13: Beck depression inventory with 13 items; RSEI: Rosenberg self-esteem inventory; BAI: Beck anxiety inventory; BHS: Beck hopelessness scale.
* $p<.001$.

13, BAI, BHS (see Table 4); and (ii) significantly and negatively correlated with the RSEI (see Table 4).

## 4. Discussion

The aim of the present study was to develop and validate a 4-item VAS version of the French CES-D Scale (Radloff, 1977) specifically designed for ESM-EMA research. The current findings indicated that both the CES-D-VAS and CES-D-VAS-VS (i.e., 4 items) represent reasonably reliable and valid measures of the depressive symptoms, applicable across clinical status (i.e., depressed vs. non-depressed individuals) and gender (i.e., male vs. female).

The primary objective of this study was to develop a very short, four items, French CES-D-VAS-VS, and to investigate its psychometric properties in a pooled sample of depressed patients and community adults. The objective underlying the development of the CES-D-VAS-VS was to reduce the time required to assess the full multidimensionality of depressive disorders in the context of extensive longitudinal or idiographic ESM-EMA studies. This was accomplished following recent recommendations on the development of short-form tests (Marsh et al., 2005; Smith et al., 2000). The fundamental requirement in developing a short form of the instrument was to start with a well-established or strong long form. Thus, we first did develop a full length VAS version of the CES-D (i.e., the CES-D-VAS) designed to assess momentary fluctuations of depressive symptoms. The obtained results showed that, in the overall sample, the postulated second-order factor model of the CES-D-VAS provided a satisfactory fit to the data and a better fit than the alternative models. These results confirm those from previous studies (Boisvert et al., 2003; Gonçalves and Fagulha, 2004; Hertzog et al., 1990; Knight et al., 1997; Rhee et al., 1999; Sheehan et al., 1995; Stommel et al., 1993) obtained with the classic Likert-type answer scale of the CES-D, suggesting that this new form of the instrument did preserve the psychometric properties of the original version. Further analyses also confirmed that the various CES-D-VAS subscales possessed internal consistency coefficients that were adequate ($\omega = .67$ to .95) and comparable to those found with the original CES-D.

The factor model obtained for the CES-D-VAS was thus used as a starting point for the development of the CES-D-VAS-VS, by selecting one item by dimension. On the basis of Marsh et al. (2005) and Smith et al. (2000) criteria, the items 12 ("*I am happy*"), 13 ("*I talk less than usual*"), 17 ("*I have crying spells or feel like it*") and 19 ("*I feel that people disliked me*") were selected. In the overall sample, results from the first-order CFA performed on those 4 items provided a satisfactory fit to the data and additional analyses confirmed that the scale presented an adequate consistency coefficients ($\omega = .80$). Furthermore, we examined the convergent validity of the selected items and of the full scale of CES-D-VAS-VS with another measure of depression and with measures of self-esteem, anxiety and hopelessness. The results showed that the 4 items and full scale scores of the CES-D-VAS-VS were moderately correlated to these measures, which concur with results from previous studies (Abramson et al., 2002; Angold et al., 1999; Gladstone and Kaslow, 1995; Gotlib and Hammens, 2002; Roberts and Monroe, 1999) and confirm the predictive and convergent validity of the newly developed instrument.

Additionally, multiple group CFA models were performed with the objective of assessing the measurement and latent mean invariance of the CES-D-VAS and of the CES-D-VAS-VS across gender and clinical groups. In the gender-based analyses, the results showed that, for both versions of the instruments, the factor loadings and intercepts of the measurement model were invariant across men and women. However, a lack of invariance was found when the errors (uniquenesses) of the items were constrained to be equal across men and women. Indeed, equality constraints across gender had to be relaxed for the uniqueness of approximately two thirds of the items from the CES-D-VAS and of three items out of four from the CES-D-VAS-VS (i.e., items 12, 17 and 19). Women tended to present a significantly higher level of measurement error on these non-invariant items. These results support those from previous studies showing that many items from the CES-D tended to present gender-related biases (Cole et al., 2000; Gelin and Zumbo, 2003; Grayson et al., 2000; Stommel et al., 1993; Yang and Jones, 2007). However, the present study indicates that these biases concern the item-specific error variances rather than their intercepts, something that had never been evaluated before. Moreover, when these biases were taken into account, the 1st-order and second-order latent means estimated on the basis of the CES-D-VAS did not differ across gender in the expected direction, although the rest of the 2nd-order measurement model did prove to be invariant across gender.

This is highly intriguing, and even worrying, since gender differences in depressive symptoms are one of the most robust findings in psychiatric epidemiology (Angold and Worthman, 1993). One logical conclusion of those results would be that those gender differences might, after all, only represent an artifact of measurement biases that have failed to be controlled in previous studies. We do not believe that such a claim can be made on the basis of this single study whose results could simply reflect sampling error or be attributed to various idiosyncrasies: (i) the reliance on a non-validated French version (this is somewhat offset by the fact that the factor validity and reliability of the CES-D-VAS appeared quite strong in the present study); and (ii) the reliance on a university-based community sample (i.e., ~50% of University students) merged with a clinically depressed sample, two types of populations in which gender differences might disappear (Nolen-Hoeksema, 1987; Scheibe et al., 2003). However, these results clearly underline the need for future studies to devote a great deal more attention to potential measurement biases in instruments designed to measure depression. Moreover, it should be noted that this is not the first time that the absence of gender based-

Author's personal copy



differences is reported on an instrument assessing the severity of depressive symptoms (rather than the prevalence of a depression diagnosis). Indeed, additional studies examining gender-based differences in depressive symptoms with various methodologies (e.g., observed score methods, latent variable models, etc.) and depression instruments (i.e., Goldberg Anxiety and Depression Scales [(Goldberg et al., 1988), Depressive experience questionnaire (Blatt et al., 1976), Beck Depression Inventory (Beck et al., 1961), and Depression scale (Salokangas et al., 1995)] are consistent with those results and reported both a lack of gender-based latent mean differences and measurement invariance (Christensen et al., 1999; Morin et al., 2010; Salokangas et al., 2002; Steer et al., 1989; Wenzel et al., 2005).

Alternatively, the results based on the very short version (i.e., the CES-D-VAS-VS) resulted in mean levels differences in depression levels that were in the expected direction and showed lower mean levels of depression in men relative to women. This is slightly intriguing given the need for short forms instruments to preserve the psychometric properties of the long forms, which is not the case, at least at the level of invariance of the mean structure of the CFA model. Conversely, the short form items were selected on the basis of multiple stringent criteria so as to provide the best representation of the latent depression construct as possible. Thus, these items could represent an optimal way of measuring depression in an ESM-EMA context, which is apparently confirmed by the fact that the CES-D-VAS-VS do present the gender-related mean level differences that depression instruments should present in order to be consistent with the results from psychiatric epidemiology. So the present results may in fact reflect either a real-absence of differences in mean-level severity across gender or not. In both cases however, both the CES-D-VAS and the CES-D-VAS-VS instrument were found to present gender-related measurement biases which should and need to be taken into account in research attempting to map gender-related mean-level differences. This clearly brings to the forefront of depression research the importance of incorporating latent variable methodologies that, for the moment, represent the only way to control for these biases.

In the next set of analyses, the invariance of the CES-D-VAS and CES-D-VAS-VS measurement models were tested across samples of clinically depressive and community subjects. For the CES-D-VAS, the results showed that although the 1st-order factor loadings proved invariant across the clinical and community subgroups, the remaining parameters of the first-order measurement model (intercepts and uniquenesses) were only partially invariant across those groups. In fact, although the invariance constraints had to be relaxed on only two intercepts, providing reasonable evidence of partial strong measurement invariance according to Meredith (1993) conception, the invariance assumption failed to hold for the great majority of the items (18/20), revealing higher levels of measurement errors in the community group. The results obtained with the CES-D-VAS-VS were highly similar in showing the non invariance of all of the items' uniquenesses, although in this model the factors loadings and items' intercepts all proved to be completely invariant across clinical/community groups. These results clearly indicate that the items from both VAS versions of this instrument have a different meaning across clinical and community subgroups. Their uses in group-based comparisons should be avoided when possible, especially when latent variables models cannot be utilized to take these biases into account. Fortunately, when latent variables models can be incorporated in the study and when these differential items biases can be taken into account, the present results showed that the latent mean differences were all in the postulated direction, revealing significantly higher levels of depression in the clinical subgroup. This being said, these results also support the measurement invariance of the higher order factor structure of the CES-D-VAS which means that – although items and lower-order factors might present some biases across subgroups – the overall depression construct measured by the CES-D-VAS appears to have the same meaning and measurement structure in those subgroups, and can thus be used with more confidence. Once again, this result should be replicated within more diversified populations and language versions. To our knowledge, this is the first time the presence of possible measurement biases have been investigated across clinical and non-clinical subgroups in depression research and, if the present results were to be replicated, they would clearly cast doubts on the ability of community-based research to generalize to clinical populations, and vice versa.

Three limitations should be kept in mind when interpreting the findings. First, this study relied exclusively on a single sample of adults and in consequence the results must be interpreted with caution. Whether the factor validity, reliability and measurement invariance of the CES-D-VAS and CES-D-VAS-VS across the overall sample and specific subgroups (i.e., gender, clinical/non clinical) can be replicated to other samples of adults, with younger or older populations or with populations from other cultural or linguistic backgrounds thus remains an open question. This clearly limits the generalizability of these results and the appropriateness of these instruments among individuals differing from those used in the present study. To ensure that this instrument could be used among adults, even from the general French population, its factor validity, reliability and measurement invariance must first be replicated in an independent and more representative sample. Similarly, the nonclinical group was rather homogeneous concerning age and social profile and consequently cannot be considered as perfectly representative of the general population. Likewise, the clinical group was solely composed of depressed individuals; although this represented an appropriate comparison group in the present study, it precludes the generalization of these findings to a more diversified clinical population. Finally, the age difference between the clinical and community groups represents a potential confounding factor. It is hard to speculate which effects this difference could have exerted on the results since the reduced size of the clinical sample precluded this verification. The effects of age should be investigated in future studies.

Second, the use of both VAS versions of the CES-D, especially the very short one, will certainly provoke a loss of screening accuracy relative to the original instrument, because it is impossible to develop an instrument that is simultaneously: (i) very short yet complete in its coverage of symptoms; (ii) sensible to momentary fluctuations yet precise in its identification of the chronic and rigid states that characterize depressive psychopathologies; and (iii) sensible to intra-individual differences in responding that are the object of ESM-EMA research yet accurate in the context of inter-

Author's personal copy



individual comparisons. However, the goal of ESM-EMA research is not screening accuracy but rather the portrayal of intra-individual idiographic processes. For these reasons, the current VAS versions should not be used for the evaluation of stable symptoms of depression or in the objective of reaching valid diagnoses. Rather, its use should be limited to ESM-EMA settings, while keeping in mind the next limitation.

Third, although the CES-D-VAS-VS was specifically designed for repeated and intensive ESM-EMA assessment in a natural setting, but was never tested in this context. In fact, there is actually no data in this study to support the usefulness of this instrument in a real ESM-EMA setting. Indeed, since such measures are generally used in an ongoing fashion (e.g., within-day, ambulatory assessments, etc.), their completion in real ESM-EMA settings will certainly involve much more chaotic and unregulated environments than the one-shot controlled conditions used in the present study. For the moment, we felt that the present study accomplished an important first step in this direction by developing such an instrument and ensuring its construct validity and its convergence with the longer and more complete versions of the CES-D before more costly ESM-EMA studies could be conducted. This goal having been accomplished, the verification of the validity and reliability of this instrument in real time and more ecologically representative contexts should be clearly addressed in future research.

## 5. Conclusion

In conclusion, the psychometric properties of the 4-item CES-D-VAS-VS were found to be adequate in French patients and community men and women. This instrument is a useful research tool in the ESM-EMA studies of depressive symptoms, since it represents the first available psychometrically valid multidimensional depression questionnaire specifically designed for this methodology. This lack of short multidimensional instruments probably explains why so many of the ESM-EMA preceding studies solely relied on mood based inventories, thus neglecting the full multidimensionality of depression. This new short multidimensional measure of depression offers the opportunity to capture a detailed picture of the dynamic and multifaceted nature of the depression. This more complete picture would greatly aid clinicians in diagnosis and treatment planning. Although traditional trait measures are still useful and important in assessing individual's perception of their own symptoms and functioning, ESM-EMA measures of depression offer a level of detail that can be essential in developing and delivering a variety of clinical interventions (Ebner-Priemer and Trull, 2009b; Solhan et al., 2009).

**Role of funding source**

This study was not financially supported.

**Conflict of interest**

None authors have conflicts of interest to disclose.

## Acknowledgments

The authors thank Valéry Balestra-Depont, Cécile Mantrant, Sylvie Gossery and Muriel Wybo for their particular contribution in the organization of the study, as well as Pablo Sillon for his help in documenting ESM-EMA methods. Completion of this study was also possible because of the dedication of the study personnel of each participating center: Clinic la Costière (Jean-Louis Just, Gérard Marro, Georges Benichou, Jean-Pierre Dany, Patrick Ferrer, Richard Moriano, Alain Salimpour, Jean-Charles Struelens), Clinic Saint-Luc (Claude Capdeville, Sophie Vincent, Christian Bessi, Pierre Bonhomme) and Hôpital de Montfavet (Laurent Layet, Marie-Noëlle Petit and Charles Dahan).

## References

Abramson, L.Y., Alloy, L.B., Hankin, B.L., Haeffel, G.J., MacCoon, D.G., Gibb, B.E., 2002. Cognitive vulnerability-stress models of depression in a self-regulatory and psychobiological context. In: Gotlib, I.H., Hammens, C.L. (Eds.), Handbook of Depression. Guilford, London (UK), pp. 268–294.

American Psychiatric Association, 1994. Diagnostic and statistical manual of mental disorders (4th ed.). Author, Washington (DC).

Angold, A., Worthman, C.W., 1993. Puberty onset of gender differences in rates of depression: a developmental, epidemiologic and neuroendocrine perspective. J. Affect. Disord. 29, 145–158.

Angold, A., Costello, E.J., Erkanli, A., 1999. Comorbidity. J. Child. Psychol. Psychiatry. 40, 57–87.

Arbuckle, J., 2006. AMOS 7.0 User's Guide. SPSS, Chicago (IL).

Barge-Schaapveld, D.Q., Nicolson, N.A., 2002. Effects of antidepressant treatment on the quality of daily life: an experience sampling study. J. Clin. Psychiatry 63, 477–485.

Barge-Schaapveld, D.Q., Nicolson, N.A., van der Hoop, R.G., De Vries, M.W., 1995. Changes in daily life experience associated with clinical improvement in depression. J. Affect. Disord. 34, 139–154.

Bebbington, P.E., 1996. The origins of sex differences in depressive disorder: bridging the gap. Int. Rev. Psychiatry 8, 295–332.

Beck, A.T., Beck, W.R., 1972. Screening depressed patients in family practice: a rapid technique. Postgrad. Med. 52, 81–85.

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

Beck, A.T., Weissman, A., Lester, D., Trexler, L., 1974. The measurement of pessimism: the hopelessness scale. J. Consult. Clin. Psychol. 42, 861–865.

Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. J. Consult. Clin. Psychol. 56, 893–897.

Beekman, A.T., van Limbeek, J., Deeg, D.J., Wouters, L., van Tilburg, W., 1994. Een screeningsinstrument voor depressie bij ouderen in de algemene bevolking: de bruikbaarheid van de Center for Epidemiologic Studies Depression Scale (CES-D) [A screening tool for depression in the elderly in the general population: the usefulness of Center for Epidemiological Studies Depression Scale (CES-D)]. Tijdschr. Gerontol. Geriatr. 25, 95–103.

Blatt, S.J., D'Afflitti, J.P., Quinlan, D.M., 1976. Experiences of depression in normal young adults. J. Abnorm. Psychol. 85, 383–389.

Boisvert, J.A., McCreary, D.R., Wright, K.D., Asmundson, G.J., 2003. Factorial validity of the center for epidemiologic studies-depression (CES-D) scale in military peacekeepers. Depress. Anxiety 17, 19–25.

Bourque, P., Beaudette, D., 1982. Étude psychométrique du questionnaire de dépression de Beck auprès d'un échantillon d'étudiants universitaires francophones [Psychometric study of the Beck Depression Inventory among a sample of French student]. Can. J. Behav. Sci. 14, 211–218.

Bouvard, M., Charles, S., Guerin, J., Aimard, G., Cottraux, J., 1992. Etude de l'échelle de désespoir de Beck (hopelessness scale). Validation et analyse factorielle [Study of Beck's hopelessness scale. Validation and factor analysis]. Encephale 18, 237–240.

Byrne, B.M., 2005. Factor analytic models: viewing the structure of an assessment instrument from three perspectives. J. Pers. Assess. 85, 17–32.

Chen, F.F., 2007. Sensitivity of goodness of fit indexes to lack of measurement. Struct. Equ. Modeling 14, 464–504.

Cheung, G.W., 2008. Testing equivalence in the structure, means, and variances of higher-order constructs with structural equation modeling. Organ. Res. Methods 11, 593–613.

Cheung, C.K., Bagley, C., 1998. Validating an American scale in Hong Kong: the Center for Epidemiological Studies Depression Scale (CES-D). J. Psychol. 132, 169–186.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of fit indexes for testing measurement invariance. Struct. Equ. Modeling 9, 233–255.

Christensen, H., Jorm, A.F., Mackinnon, A.J., Korten, A.E., Jacomb, P.A., Henderson, A.S., Rodgers, B., 1999. Age differences in depression and

Author's personal copy



anxiety symptoms: a structural equation modelling analysis of data from a general population sample. Psychol. Med. 29, 325–339.
Cole, S.R., Kawachi, I., Maller, S.J., Berkman, L.F., 2000. Test of item-response bias in the CES-D scale. experience from the New Haven EPESE study. J. Clin. Epidemiol. 53, 285–289.
Collet, L., Cottraux, J., 1986. Inventaire abrégé de la dépression de Beck (13 items). Etude de la validité concurrente avec les échelles de Hamilton et de ralentissement de Widlocher [The shortened Beck depression inventory (13 items). Study of the concurrent validity with the Hamilton scale and Widlocher's retardation scale]. Encephale 12, 77–79.
Dershem, L.D., Patsiorkovski, V.V., O'Brien, D.J., 1996. The use of the CES-D for measuring symptoms of depression in three rural Russian villages. Soc. Indic. Res. 39, 89–108.
Ebner-Priemer, U.W., Trull, T.J., 2009a. Ambulatory assessment: an innovative and promising approach for clinical psychology. Eur. Psychol. 14, 109–119.
Ebner-Priemer, U.W., Trull, T.J., 2009b. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol. Assess. 21, 463–475.
Fava, G.A., 1983. Assessing depressive symptoms across cultures: Italian validation of the CES-D self-rating scale. J. Clin. Psychol. 39, 249–251.
Flett, G.L., Vredenburg, K., Krames, L., 1997. The continuity of depression in clinical and nonclinical samples. Psychol. Bull. 121, 395–416.
Fountoulakis, K., Iacovides, A., Kleanthous, S., Samolis, S., Kaprinis, S.G., Sitzoglou, K., St Kaprinis, G., Bech, P., 2001. Reliability, validity and psychometric properties of the Greek translation of the Center for Epidemiological Studies-Depression (CES-D) Scale. BMC Psychiatry 1, 3.
Freeston, M.H., Ladouceur, R., Thibodeau, N., Gagnon, F., Rheaume, J., 1994. The Beck anxiety inventory. Psychometric properties of a French translation. Encephale 20, 47–55.
Führer, R., Rouillon, F., 1989. La version française de l'échelle CES-D. Description et traduction de l'échelle d'autoévaluation [The French version of CES-D: Description and translation of the self-report scale]. Psychiatr. Psychobiol. 4, 163–166.
Gelin, M.N., Zumbo, B.D., 2003. Differential item functioning results may change depending on how an item is cored: an illustration with the center for epidemiologic studies depression scale. Educ. Psychol. Meas. 63, 65–74.
Gladstone, T.R., Kaslow, N.J., 1995. Depression and attributions in children and adolescents: a meta-analytic review. J. Abnorm. Child Psychol. 23, 597–606.
Goldberg, D., Bridges, K., Duncan-Jones, P., Grayson, D., 1988. Detecting anxiety and depression in general medical settings. Br. Med. J. 297, 897–899.
Golier, J.A., Yehuda, R., Schmeidler, J., Siever, L.J., 2001. Variability and severity of depression and anxiety in post traumatic stress disorder and major depressive disorder. Depress. Anxiety 13, 97–100.
Gonçalves, B., Fagulha, T., 2004. The Portuguese Version of the Center for Epidemiologic Studies Depression Scale (CES-D). Eur. J. Psychol. Assess. 20, 339–348.
Gordijn, M.C., Beersma, D.G., Bouhuys, A.L., Reinink, E., Van den Hoofdakker, R.H., 1994. A longitudinal study of diurnal mood variation in depression; characteristics and significance. J. Affect. Disord. 31, 261–273.
Gotlib, I.H., Hammens, C.L., 2002. Handbook of depression. The Guilford press, New-York (NY).
Grayson, D.A., Mackinnon, A., Jorm, A.F., Creasey, H., Broe, G.A., 2000. Item bias in the Center for Epidemiologic Studies Depression Scale: effects of physical disorders and disability in an elderly community sample. J. Gerontol. B Psychol. Sci. Soc. Sci. 55, 273–282.
Hankin, B.L., Abramson, L.Y., 1999. Development of gender differences in depression: description and possible explanations. Ann. Med. 31, 372–379.
Hankin, B.L., Fraley, R.C., Lahey, B.B., Waldman, I.D., 2005. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. J. Abnorm. Psychol. 114, 96–110.
Hautzinger, M., 1988. The CES-D scale: a depression-rating scale for research in the general population. Diagnostica 34, 167–173.
Hertzog, C., Alstine, J.V., Usala, P.D., Hultsh, D.F., Dixon, R., 1990. Measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in older populations. Psychol. Assess. 2, 64–72.
Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Modeling 6, 1–55.
Hyde, J.S., Mezulis, A.H., Abramson, L.Y., 2008. The ABCs of depression: integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. Psychol. Rev. 115, 291–313.
Judd, L.L., 1997. Pleomorphic expressions of unipolar depressive disease: summary of the 1996 CINP President's Workshop. J. Affect. Disord. 45, 109–116.
Knight, R.G., Williams, S., McGee, R., Olaman, S., 1997. Psychometric properties of the Centre for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. Behav. Res. Ther. 35, 373–380.
Larson, R., Csikszentmihalyi, M., 1983. The experience sampling method. New Dir. Meth. Soc. Behav. Sci. 15, 41–56.
Leach, L.S., Christensen, H., Mackinnon, A.J., 2008. Gender differences in the endorsement of symptoms for depression and anxiety: are gender-biased items responsible? J. Nerv. Ment. Dis. 196, 128–135.
Lecrubier, Y., Sheehan, D.V., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., Janavs, J., Dunbar, G.C., 1997. The MINI international neuropsychiatric interview (M.I.N.I.) a short diagnostic structured interview: reliability and validity according to the CIDI. Eur. Psychiatry 12, 224–231.
MacCallum, R.C., Browne, M.W., Sugawara, H.M., 1996. Power analysis and determination of sample size for covariance structure modeling. Psychol. Meth. 1, 130–149.
Marsh, H.W., Ellis, L.A., Parada, R.H., Richards, G., Heubeck, B.G., 2005. A short version of the Self Description Questionnaire II: operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. Psychol. Assess. 17, 81–102.
McCormack, H.M., Horne, D.J., Sheather, S., 1988. Clinical applications of visual analogue scales: a critical review. Psychol. Med. 18, 1007–1019.
McDonald, R.P., 1999. Test theory: A unified treatment. Lawrence Erlbaum Associates Inc., Mahwah (NJ).
Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58, 525–543.
Meredith, W., Teresi, J.A., 2006. An essay on measurement and factorial invariance. Med. Care 44, S69–S77.
Mitchell, A.J., Coyne, J.C., 2007. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br. J. Gen. Pract. 57, 144–151.
Morin, A.J.S., Janosz, M., Larivée, S., 2010. The Montreal Adolescent Depression Development Project (MADDP): School life and depression following high school transition. Psychiatr. Res. J. 3, 1–50.
Moskowitz, D., Russell, J.J., Sadikaj, G., Sutton, R., 2009. Measuring people intensively. Can. Psychol. 50, 131–140.
Myin-Germeys, I., Peeters, F., Havermans, R., Nicolson, N.A., DeVries, M.W., Delespaul, P., Van Os, J., 2003. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. Acta Psychiatr. Scand. 107, 124–131.
Nolen-Hoeksema, S., 1987. Sex differences in unipolar depression: evidence and theory. Psychol. Bull. 101, 259–282.
Peeters, F., Nicolson, N.A., Berkhof, J., Delespaul, P., deVries, M., 2003. Effects of daily events on mood states in major depressive disorder. J. Abnorm. Psychol. 112, 203–211.
Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., Nicolson, N.A., 2006. Diurnal mood variation in major depressive disorder. Emotion 6, 383–391.
Perreira, K., Deeb-Sossa, N., Harris, K., Bollen, K., 2005. What are we measuring? An evaluation of The CES-D across race/ethnicity and immigrant generation. Soc. Forces 83, 1567–1602.
Piccinelli, M., Wilkinson, G., 2000. Gender differences in depression. Critical review. Br. J. Psychiatry 177, 486–492.
Pickles, A., Angold, A., 2003. Natural categories or fundamental dimensions: on carving nature at the joints and the rearticulation of psychopathology. Dev. Psychopathol. 15, 529–551.
Putnam, K.M., McSweeney, L.B., 2008. Depressive symptoms and baseline prefrontal EEG alpha activity: a study utilizing Ecological Momentary Assessment. Biol. Psychol. 77, 237–240.
Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1, 385–401.
Rhee, S.H., Petroski, G.F., Parker, J.C., Smarr, K.L., Wright, G.E., Multon, K.D., Buchholz, J.L., Komatireddy, G.R., 1999. A confirmatory factor analysis of the Center for Epidemiologic Studies Depression Scale in rheumatoid arthritis patients: additional evidence for a four-factor model. Arthritis Care Res. 12, 392–400.
Roberts, J.E., Monroe, S.M., 1999. Vulnerable self-esteem and social processes in depression: toward an interpersonal model of self-esteem regulation. In: Joiner Jr., T.E., Coyne, J.C. (Eds.), The interactional nature of depression: advances in interpersonal approaches. American Psychological Association, Washington (DC), pp. 149–187.
Rosenberg, M., 1965. Society and the adolescent self-image. Princeton University Press, Princeton (NJ).
Salokangas, R.K., Poutanen, O., Stengard, E., 1995. Screening for depression in primary care. Development and validation of the Depression Scale, a screening instrument for depression. Acta Psychiatr. Scand. 92, 10–16.
Salokangas, R.K., Vaahtera, K., Pacriev, S., Sohlman, B., Lehtinen, V., 2002. Gender differences in depressive symptoms. An artefact caused by measurement instruments? J. Affect. Disord. 68, 215–220.

Author's personal copy



Scheibe, S., Preuschhof, C., Cristi, C., Bagby, R.M., 2003. Are there gender differences in major depression and its response to antidepressants? J. Affect. Disord. 75, 223–235.
Shafer, A.B., 2006. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. J. Clin. Psychol. 62, 123–146.
Sheehan, T.J., Fifield, J., Reisine, S., Tennen, H., 1995. The measurement structure of the Center for Epidemiologic Studies Depression Scale. J. Pers. Assess. 64, 507–521.
Sheehan, D.V., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M.F., Dunbar, G.C., 1997. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. Eur. Psychiatry 12, 232–241.
Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl 20), 22–33 quiz 34–57.
Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. Annu. Rev. Clin. Psychol. 4, 1–32.
Smith, G.T., McCarthy, D.M., Anderson, K.G., 2000. On the sins of short-form development. Psychol. Assess. 12, 102–111.
Solhan, M.B., Trull, T.J., Jahng, S., Wood, P.K., 2009. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol. Assess. 21, 425–436.
Sprock, J., Yoder, C.Y., 1997. Women and Depression: An Update on the Report of the APA Task Force. Sex Roles 36, 269–303.
Steer, R.A., Beck, A.T., Brown, G., 1989. Sex differences on the revised Beck Depression Inventory for outpatients with affective disorders. J. Pers. Assess. 53, 693–702.
Stommel, M., Given, B.A., Given, C.W., Kalaian, H.A., Schulz, R., McCorkle, R., 1993. Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). Psychiatry Res. 49, 239–250.
Stone, A.A., Shiffman, S., 1994. Ecological momentary assessment (EMA) in behavioral medicine. Ann. Behav. Med. 16, 199–202.
Thiele, C., Laireiter, A.R., Baumann, U., 2002. Diaries in clinical psychology and psychotherapy: a selective review. Clin. Psychol. Psychother. 9, 1–37.
Vallières, E.F., Vallérand, R., 1990. Traduction et validation canadienne-française de l'échelle de l'estime de soi de Rosenberg [French-Canadian translation and validation of the Rosenberg's Self-Esteem Scale]. Int. J. Psychol. 25, 305–316.
Vandenberg, R.J., Lance, C.E., 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ. Res. Methods 3, 4–70.
Vazquez, F.L., Blanco, V., Lopez, M., 2007. An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. Psychiatry Res. 149, 247–252.
Wenzel, A., Steer, R.A., Beck, A.T., 2005. Are there any gender differences in frequency of self-reported somatic symptoms of depression? J. Affect. Disord. 89, 177–181.
Wolk, S., Weissman, M., 1995. Women and depression: an up-date. Rev. Psychiatry. 14, 227–259.
World Health Organization, 2000. International Statistical Classification of Diseases and Related Health Problems (10th ed.). Author, Geneva (CH).
Yang, F.M., Jones, R.N., 2007. Center for Epidemiologic Studies-Depression Scale (CES-D) item response bias found with Mantel-Haenszel method was successfully replicated using latent variable modeling. J. Clin. Epidemiol. 60, 1195–1200.