



UNIVERSIDADE D  
COIMBRA

Ana Elisabete Oliveira Tomás

**PRELIMINARY VALIDATION STUDY AND  
PSYCHOMETRIC PROPERTIES OF THE  
MONTGOMERY-ASBERG DEPRESSION SCALE  
EUROPEAN PORTUGUESE PATIENT'S SELF-REPORT  
VERSION**

**Dissertação no âmbito do Mestrado em Intervenções  
Cognitivo-Comportamentais em Psicologia Clínica e da Saúde orientada  
pela Professora Doutora Maria Cristina Cruz Sousa Portocarrero Canavarro e  
Doutora Ana Ganho Ávila Costa e apresentada à Faculdade de  
Psicologia e de Ciências da Educação da Universidade de Coimbra.**

Julho de 2022

Faculdade de Psicologia e de Ciências da Educação  
da Universidade de Coimbra

PRELIMINARY VALIDATION STUDY AND  
PSYCHOMETRIC PROPERTIES OF  
THE MONTGOMERY-ASBERG DEPRESSION  
SCALE:  
EUROPEAN PORTUGUESE PATIENT'S SELF-  
REPORT VERSION

Ana Elisabete Oliveira Tomás

Dissertação de Mestrado em Intervenções Cognitivo-Comportamentais em Psicologia Clínica e da Saúde orientada pela Professora Doutora Maria Cristina Cruz Sousa Portocarrero Canavarro e Doutora Ana Ganho Ávila Costa e apresentada à Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.

Julho de 2022



UNIVERSIDADE D  
COIMBRA

### **Declaração de integridade**

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

### **Statement of integrity**

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

### **Enquadramento institucional**

A presente dissertação foi realizada no âmbito do projeto estratégico do Centro de Investigação em Neuropsicologia e Intervenção Cognitivo-Comportamental (CINEICC) (UIDB/00730/2020).

### **Institutional framework**

The present dissertation was developed within the strategic project of the Center for Research in Neuropsychology and Cognitive-Behavioral Intervention (CINEICC) (UIDB/00730/2020).



## Acknowledgments

---

À Professora Doutora Maria Cristina Canavarro, pelo profissionalismo e preocupação com que conduz sempre o seu trabalho.

À Doutora Ana Ganho Ávila, pela incansável dedicação durante todo este percurso. Agradeço toda a ajuda que me ofereceu e o cuidado que teve nos momentos mais difíceis.

Ao meu pai, pelo apoio incondicional que me ofereceu ao longo de toda a minha vida, pelas horas infinitas que me ouviu atentamente a praticar as minhas apresentações para trabalhos académicos e por todos os conselhos que me oferece.

Aos meus avós, Diamantino e Aninhas, por serem figuras constantes na minha vida que me apoiam em todas as minhas decisões.

À minha avó Palmira, pela longas conversas e por me encorajar a seguir em frente.

Ao meu avô Carlos, pelo amor que me deu enquanto estive neste mundo e pelo amor que continuo a sentir embora já não esteja cá.

À minha mãe, pelo orgulho que tem em mim.

À minha tia Maria João, por todos os momentos que passamos juntas.

Ao meu primo Luís, por estar presente em ocasiões importantes.

À minhas amigas, Inês, Diana, Patrícia e Joana, pela amizade que formamos em Coimbra e as memórias que levo agora comigo.

À minha afilhada, Patrícia, por me escolher para a acompanhar neste percurso académico.

A todas as pessoas que participaram neste projeto, pelo valioso contributo que disponibilizaram.

## Resumo

**Objetivos.** As perturbações depressivas contribuem de forma substancial para a carga global de doença com a Perturbação Depressiva *Major* (PDM) a atingir na Europa uma prevalência ao longo da vida de 11,32%. O principal objetivo deste estudo foi examinar a validade e confiabilidade da versão de autorrelato da Escala de Avaliação de Depressão de Montgomery-Asberg (MADRS-S) para a população de língua Portuguesa Europeia e ainda avaliar a sua invariância de género e confiabilidade teste-reteste. **Métodos.** Uma amostra de 204 participantes (84.8% mulheres) com idade média de 31.07 anos ( $DP = 12.81$ ) completou um protocolo *online* constituído pelas versões em Português Europeu da MADRS-S, Inventário de Estado-Traço de Ansiedade (STAI Y) e Inventário de Depressão de Beck-II (BDI-II). Um subconjunto de 30 participantes respondeu ao mesmo protocolo *online* com um intervalo de 1 mês. Uma análise ROC foi efetuada para determinar o ponto de corte entre participantes deprimidos e não deprimidos. **Resultados.** A unidimensionalidade do modelo foi confirmada por análise fatorial confirmatória (AFC). O modelo de um fator apresentou boa consistência interna (.88) e excelente estabilidade temporal (coeficiente de correlação intraclasse de .95). Não foi possível estabelecer invariância entre os géneros. O ponto de corte de 11 mostrou uma sensibilidade de 86% e especificidade de 86.5% (área sob a curva [AUC] = .94). **Conclusões.** A versão em Português Europeu da MADRS-S apresenta propriedades psicométricas satisfatórias e pode ser uma medida útil para avaliar sintomatologia depressiva em contextos clínicos e de investigação. Devido às diferenças de género encontradas na estrutura da escala, apenas recomendamos a utilização da MADRS-S em mulheres.

*Palavras-chave:* depressão, MADRS-S, análise fatorial confirmatória, validade, confiabilidade

## Abstract

---

**Objectives.** Depressive disorders contribute substantially to the global burden of disease with Major Depressive Disorder (MDD) reaching a lifetime prevalence of 11.32% in Europe. The main goal of this study was to examine the validity and reliability of the self-rated version of the Montgomery-Asberg Depression Rating Scale (MADRS-S) for a European Portuguese-speaking population and also assess its invariance across gender and test-retest reliability. **Methods.** A sample of 204 participants (84.8% women) with a mean age of 31.07 years ( $SD = 12.81$ ) completed an online protocol that included the European Portuguese versions of the MADRS-S, State-Trait Anxiety Inventory (STAI Y), and Beck Depression Inventory-II (BDI-II). A subset of 30 participants answered the same online protocol with a 1-month interval. A ROC analysis was performed to determine a cut-off point between depressed and non-depressed participants. **Results.** The unidimensionality of the model was confirmed by confirmatory factor analysis (CFA). The one-factor model showed good internal consistency (.88) and excellent temporal stability (intraclass correlation coefficient of .95). Invariance across gender could not be established. A cut-off score of 11 showed a sensitivity of 86% and specificity of 86.5% (area under the curve [AUC] = .94). **Conclusions.** The European Portuguese version of the MADRS-S shows satisfactory psychometric properties and can be a useful measure of depressive symptoms in clinical and research settings. Due to gender differences in the structure of the scale, we only recommend the use of MADRS-S in women.

*Keywords:* depression, MADRS-S, confirmatory factor analysis, validity, reliability

## Table of Contents

---

Acknowledgments.....	4
Resumo .....	5
Abstract .....	6
Table of Contents .....	7
Preliminary Validation Study And Psychometric Properties Of The Montgomery-Asberg Depression Scale: European Portuguese Patient’s Self-Report Version.....	8
Methods .....	12
Results .....	19
Discussion .....	22
References .....	26



## **Preliminary Validation Study And Psychometric Properties Of The Montgomery-Asberg Depression Scale: European Portuguese Patient's Self-Report Version**

---

In 2019, mental disorders were one of the main contributors to the global burden of disease with depressive disorders accounting for 37.3% of years of healthy life lost due to illness (GBD 2019 Mental Disorders Collaborators, 2022).

According to the Diagnostic and Statistical Manual of Mental Disorders (5<sup>th</sup> ed; DSM-5; American Psychiatric Association [APA], 2013), Major Depressive Disorder (MDD) is characterized by changes in several aspects of life (e.g., mood, appetite, sleep, and concentration) interfering with normal functioning when present almost every day for at least two weeks. In a recent systematic review comparing prevalence by continent (Gutiérrez-Rojas et al., 2020), Europe reported the highest numbers of MDD with a lifetime prevalence of 11.32% and a 12-month prevalence of 5.2%.

The clinician-rated (Montgomery & Asberg, 1979) and the self-rated (Svanborg & Asberg, 1994) versions of the Montgomery-Asberg Depression Rating Scale (MADRS) have been shown to have good correlations to the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960; Heo et al., 2007) and the Beck Depression Inventory-II (BDI-II; Beck et al., 1996; Wikberg et al., 2015), respectively. The HDRS and the BDI-II are considered gold standard rating scales for depression (Cusin et al., 2010).

The MADRS (clinician-rated version) was originally proposed due to the lack of rating scales for depression that would be sensitive to change over time and that could assess the efficacy of pharmacological treatment in clinical trials. MADRS contains 10 items extracted from the Comprehensive Psychopathological Rating Scale (CPRS; Asberg et al., 1978) that include apparent sadness, reported sadness, inner tension, sleep, appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts.

The MADRS-S (self-rated version) was later developed and includes nine items: mood, feelings of unease, sleep, appetite, ability to concentrate, initiative, emotional involvement, pessimism, and zest for life. The difference concerning the number of items between both scales is due to the suppression of the “apparent sadness” item of MADRS given its clinician-rated nature.

In MADRS-S, participants are instructed to respond based on their subjective experience during the last three days. Each multiple-choice item is rated on a seven-point scale with scores ranging from 0 (e.g., “I have no difficulty in concentrating”) to 6 (e.g., “I am quite

unable to concentrate on anything at all”). The total score is estimated by the absolute sum of the items, ranging from 0 to 54 points with higher scores indicating greater degrees of severity.

Svanborg and Asberg (1994) demonstrated in two different studies that MADRS-S has a good correlation with the clinician-rated version (.80 and .94). Subsequent studies of the psychometric properties of the MADRS-S have yielded satisfactory results. Fantino and Moore (2009) observed a sample of 278 adult outpatients diagnosed with MDD and demonstrated good internal consistency (.84), satisfactory test-retest reliability (.78), and good sensitivity to change with an effect size of 2.8 between baseline and after eight weeks of pharmacological treatment. Moreover, the factor analysis conducted demonstrated the one-factor structure of the paper-pencil version of the MADRS-S. A consequent study examining the psychometric properties of an online version of the MADRS-S showed no significant differences with the paper-pencil version and indicated a good correlation between the two formats (.84) (Holländare et al., 2010). Recently, MADRS-S psychometric properties have also been assessed for adolescent psychiatric patients, showing good internal consistency (.87) and satisfactory accuracy (Ntini et al., 2020). Here, a two-factor model showed good adjustment. The first factor included the items concerning mood, feelings of unease, appetite, emotional involvement, pessimism, and zest for life while the second factor included the items on sleep, ability to concentrate, and initiative. MADRS-S has also been used to assess the efficacy of pharmacological (e.g., Jain et al., 2013) and psychological treatments (e.g., Westas et al., 2022) due to its sensitivity to change.

When comparing clinician-rated and self-rated measures of depression for outcomes of pharmacological treatments for MDD, Uher et al., (2012) suggested that both offer valuable information and can complement each other. This recommendation has been substantiated by other studies. For example, when examining depressive symptomology and suicidal ideation in the context of pharmacological treatment for depression, self-rated improvement in BDI-II items assessing subjective experiences (e.g., sadness, fatigue) were the best predictors for a decrease in suicidal ideation (Keilp et al., 2018). Adding some of the items assessed by the HDRS (clinician-rated; e.g., depressed mood, guilt) helped to enhance that prediction and reinforced the need for both types of measures.

Self-rated measures can also be useful in the clinical setting by encouraging patients to reflect on their condition, feel more comfortable introducing topics that have not yet been discussed, and informing patients of the clinician’s interest in their subjective experience (Greenhalgh et al., 2018). In the case of the MADRS-S, Wikberg et al. (2016) demonstrated that patients’ experience using MADRS-S in a primary care setting with a general practitioner Preliminary Validation Study and Psychometric Properties of the Montgomery-Asberg Depression Scale: European Portuguese Patient’s Self-report Version 9

was overall positive. MADRS-S offered patients confirmation concerning the existence and severity of their depressive symptoms, reassured them that the general practitioner was accounting for their difficulties, and clarified the reasons supporting the need for treatment. Patients also valued the completion of MADRS-S itself rather than its score. That is, the completion of the scale was experienced by patients as a manifestation of the clinician's interest in their requests for health care. At the primary care level, MADRS-S has been considered a faster measure of depression when compared to the BDI-II due to the reduced number of items (21 vs. 9; Wikberg et al., 2015). Moreover, contrary to BDI-II, MADRS-S is in the public domain as a free-access tool.

Public health organizations encourage the use of self-rated measures and outcomes. In a report presented by the Health Evidence Network (HEN), 1485 indicators were identified in the evaluation of the performance of health systems in the World Health Organization (WHO) European Region (Fekri et al., 2018). Among the indicators used for the assessment of improved health, patient-reported outcome measures (PROMs) were present and represented a change to the concept of health outcomes. Indicators such as quality of life were considered to go beyond more classical indicators such as mortality. The inclusion of PROMs was recommended for future health policies to better represent patients' subjective experiences.

MADRS-S was so far translated and validated for a French-speaking population in a sample of 63 adult outpatients diagnosed with several affect disorders (Bondolfi et al., 2010). The French version of the MADRS-S exhibited good internal consistency on two different occasions (.85 and .94). A principal component analysis determined that one single factor explained a substantial portion of the variance (68.8). MADRS-S has also been adapted for a Malay-speaking population using a sample of 50 individuals diagnosed with depression and 100 individuals without a diagnosis (Yee et al., 2015). The Malay version demonstrated good internal consistency (.78) and good test-retest reliability (.88). A principal component analysis determined that one single factor accounted for a sizable portion of the variance (61.3). MADRS-S showed cross-cultural validity and reliability, and a one-factor structure has been established.

With the current study, we aim to offer the European Portuguese-speaking population an important access free and self-rated measure of depressive symptoms. At the clinical level, we intend to reinforce the already available traditional depression evaluation tools and offer a more rapid and free-access tool. Additionally, at the research level, the European Portuguese version of the MADRS-S will allow cross-cultural studies using this scale.

To our knowledge, this is the first study to conduct a confirmatory factor analysis (CFA) of the European Portuguese version of MADRS-S. Additionally, we will assess MADRS-S construct validity (to ensure the scale accurately represents the construct it plans to measure) and construct reliability (to confirm that the scores obtained are consistent when repeated under the same conditions). Moreover, we will measure its invariance across gender (to evaluate whether the construct holds the same psychometric characteristics in both men and women). Finally, we will perform a Receiver Operating Characteristic (ROC) Curve analysis to suggest a cutoff score, so diagnostic accuracy for the European Portuguese population can be established.

## Methods

### Participants

Participants were recruited between March and July of 2021 through public announcements on social media platforms available to the general population and during classes to university students. Inclusion criteria included: 1) being 18 years of age or older; 2) having internet access; 3) being proficient in the European Portuguese language; 4) consenting to participate in the study. No exclusion criteria were applied. In total, 208 participants completed the online survey using the data collection website LimeSurvey®. Four participants were excluded given their indication of a non-Portuguese nationality.

The final sample included 204 participants (84.8% women, 14.7% men and 0.5% non-binary) with a mean age of 31.07 years ( $SD = 12.81$ ) (see **Table 1**).

**Table 1**

*Descriptive statistics of the sample*

		<i>N</i> = 204
Age, years <sup>a</sup>		
	<i>M</i>	31.07
	<i>SD</i>	12.81
Sex		
	Men (%)	30 (14.7)
	Women (%)	174 (84.8)
	Non-Binary (%)	1 (0.5)
Marital status		
	Single (%)	136 (66.7)
	Married/cohabiting (%)	57 (27.9)
	Divorced (%)	8 (3.9)
	Widow (%)	3 (1.5)
Area of residence		
	Rural (%)	70 (34.3)
	Urban (%)	134 (65.7)
Education, years		
	<i>M</i>	14.54

	<i>SD</i>	3.87
Academic qualification		
	Primary	1 (0.5)
	Secondary	80 (39.2)
	University	123 (60.2)
Profession		
	Employed/Student	198 (97)
	Unemployed	6 (3.0)
Chronic illness		
	Yes	36 (17.6)
	No	168 (82.4)
Clinical Diagnosis (past or present)		
	Referred by Clinician	46 (22.5)
	Self-report	116 (56.9)
Family member with depression (past or present)		
	Yes (%)	133 (65.2)
	No (%)	71 (34.8)
Psychiatric/psychological treatment (past or present)		
	Yes (%)	71 (34.8)
	No (%)	133 (65.2)
Duration of psychiatric/psychological treatment (past or present), months		
	<i>M</i>	18.86
	<i>SD</i>	22.05

---

<sup>a</sup>One participant did not insert a valid age

## Procedure

Authorization for this study was granted by the local ethics committee. Participants were requested for consent after receiving information about the study and what their involvement would require. They did not receive any compensation (monetary or other) for their participation. The protocol presented to all participants included online versions of an in-

house sociodemographic and clinical questionnaire and the European Portuguese versions of MADRS-S, State-Trait Anxiety Inventory (STAI Y), and BDI-II. Measures were presented to participants in the same order.

## **Measures**

### ***Sociodemographic and Clinical Questionnaire***

We used an in-house self-report questionnaire to gather sociodemographic data regarding gender, age, nationality, area of residence, civil status, academic qualification, number of years of schooling completed, and profession. We also developed an in-house self-report questionnaire to collect clinical data concerning chronic illnesses, previous or current diagnosis of depression, previous or current self-assessed experience of depression, diagnosis of depression in a family member, and previous or current psychiatric/psychological treatment.

### ***Montgomery-Asberg Depression Rating Scale (self-rated)***

The self-report version of the Montgomery-Asberg Depression Rating Scale (MADRS-S; Svanborg & Asberg, 1994) is a self-assessment measurement for depression that includes nine multiple-choice items with half steps (e.g., “Here you should assess your interest in your surroundings, in other people, and in activities that normally give you pleasure”) rated on a seven-point scale in which responses range from 0 (e.g., “I am interested and involved in my surroundings, and this gives me pleasure”) to 6 (“I no longer have any feelings. I feel painfully indifferent, even toward those closest to me”). Total scores vary from 0 to 54 points and are obtained through a sum of the absolute score on each item. Higher scores indicate a higher degree of depressive symptomology severity. We followed the International Test Commission (2005) guidelines for test development and adaptation to reach the European Portuguese version of the MADRS-S. We translated the original version into European Portuguese and performed a back-translation to English for semantic equivalence. The translation was conducted independently by two junior clinical psychologists, and the final version was reached by a senior mental health researcher, all Portuguese language native speakers. Back-translation was conducted by a Portuguese language native speaker proficient in the English language. A face validity test was completed with three native speakers (of distinctive backgrounds and education levels). Small changes were made regarding wording, but no major alterations were performed in its structure, content, and number of items.

### ***State-Trait Anxiety Inventory***

The European Portuguese version of the State-Trait Anxiety Inventory (STAI; Silva, 2003; Spielberger et al., 1983) was used to measure trait anxiety. STAI is a self-assessment instrument comprised of two scales that measure state (Y-1) and trait anxiety (Y-2). The protocol for the current study incorporated only the scale for trait anxiety (Y-2) because we were not interested in momentary anxiety symptoms but rather in stable anxiety levels. Trait anxiety refers to stable individual differences in the predisposition to experience anxiety and consequently to also experience more frequent and intense levels of state anxiety. The STAI Y-2 contains 20 items (e.g., “I feel calm”) which are scored on a four-point Likert scale ranging from 1 (“Not at all”) to 4 (“Very much”). The total score ranges from 20 to 80 points and is obtained through the absolute sum of the items (items 21, 23, 26, 27, 30, 33, 34, 36, and 39 are inverted). Higher total scores indicate higher levels of trait anxiety. The European Portuguese version of the STAI showed good internal consistency on both scales (Silva, 2003). More specifically, the scale for trait anxiety (Y-2) revealed satisfactory internal consistency for men (.89) and women (.89). In the present study, Cronbach’s alpha value was .95.

### ***Beck Depression Inventory-II***

The European Portuguese version of the Beck Depression Inventory-II (BDI-II; (Beck et al., 1996; Campos & Gonçalves, 2011) is a self-reported measurement for depression. It contains 21 items (e.g., “Loss of Energy”) which are answered on a multiple-choice four-point scale (or a seven-point scale for items 16 and 18) ranging from 0 (e.g., “I have as much energy as ever”) to 3 (e.g., “I don't have enough energy to do anything”). The total score ranges from 0 to 63 points with higher scores suggesting a greater degree of severity of depressive symptomology. According to Beck et al., (1996), scores between 0 and 13 points indicate “minimal” depression, between 14 and 19 points “mild”, between 20-28 points “moderate”, and between 29 and 63 points “severe”. The European Portuguese version of the BDI-II showed good internal consistency with a Cronbach’s alpha value of .91 (Campos & Gonçalves, 2011). In the present study, Cronbach’s alpha value was .93.

### **Statistical Analysis**

MADRS-S factorial structure, reliability, and validity analysis were conducted using the *Statistical Package for the Social Sciences* (SPSS; v25.0) and *Mplus* (v8.3; Muthén & Muthén, 1998-2017).



### ***Factorial Structure and Measurement Invariance***

We performed a CFA using *Mplus* to determine the factorial structure of the MADRS-S. Considering previous studies through which a unidimensional model was confirmed (e.g., Fantino & Moore, 2009), we estimated a unidimensional model with all items loading onto one single factor representing the latent construct of depression. Because the data followed a non-normal distribution, we used the maximum likelihood method with robust standard errors (MLR) as an estimator (Muthén & Muthén, 1998-2017). For assessment of model fit, the chi-square statistic ( $\chi^2$ ), the Root Mean Squared Error of Approximation (RMSEA), and the Comparative Fit Index (CFI) were considered goodness-of-fit indexes (Browne & Cudeck, 1992). We considered the fit index thresholds recommended by Hair et al., (2010) that suggest an acceptable model fit when  $\chi^2$  is not significant, RMSEA < .08, and CFI  $\geq$  .97. Regarding RMSEA, Browne and Cudeck (1992) deemed values of .05 or below a good fit, between .05 and .08 an adequate fit, values between .08 and .10 a mediocre fit, and any values above 1.0 a poor fit.

We evaluated measurement invariance across gender through an analysis of the configural, metric (weak), and scalar (strong) invariance of the model (Chen, 2007). We followed Dimitrov (2010) according to which configural invariance should be first tested by applying the same one-factor model to each group, separately. Then configural invariance should be tested in the whole sample by performing a multi-group CFA where the one-factor model is applied without any constraints to men and women simultaneously (van de Schoot et al., 2012). The baseline model for further invariance testing is this configural model. Metric invariance is tested by constraining factor loadings to be equal across groups and comparing the fit of the metric model to the fit of the configural model using the  $\chi^2$  difference test. If the difference between the models is significant at a .05 level of significance, it means a significant worsening of the model when constraints are applied, and full metric invariance should not be established (Dimitrov, 2010). If metric invariance is not confirmed, partial metric invariance should be tested by sequentially removing the constraints applied to the factor loadings, allowing them to vary between groups. Upon establishing (partial) metric invariance, maintaining the constraints used on the metric model, scalar invariance should be tested by constraining items' intercepts to be equal across groups. Because measurement invariance tests are a series of hierarchically nested CFAs, whenever invariance is not established at one higher level, we should not pursue the next one (Putnick & Bornstein, 2016).

### ***Reliability***

We calculated MADRS-S internal consistency by using Cronbach's alpha to assess items' correlations (Cronbach, 1951). Additionally, we assessed composite reliability (CR) which considers the factor loadings calculated in the CFA (Bacon et al., 1995). We estimated the average variance extracted (AVE) to determine the amount of variance shared by the items that can be explained by the proposed latent construct of depression (as opposed to the amount of variance that can be attributed to measurement error) (Fornell & Larcker, 1981). Using data from a subset of participants that answered MADRS-S twice with a 1-month interval, we estimated stability over time using the intraclass correlation coefficient (ICC) (Bland & Altman, 2003). Following the guidelines of Koo and Li (2016), values below .50 indicated poor reliability, between .50 and .75 indicated moderate reliability, between .75 and .90 indicated good reliability, and above .90 indicated excellent reliability. For Cronbach's alpha, values above .70 were deemed adequate (Kline, 2011). According to Hair et al., (2013), AVE should be .50 or higher and CR should be .70 or higher.

### ***Validity***

We studied floor and ceiling effects to ensure that a considerable percentage of participants did not obtain either minimum or maximum scores on the scale. We considered values below 15% to be adequate (Terwee et al., 2007). Then, we assessed concurrent and discriminant validity. We estimated concurrent validity by determining if MADRS-S had a strong relationship with another measure of depression – the BDI-II. We correlated the total scores of the MADRS-S and the BDI-II using Pearson product-moment correlation. We expected a large and significant correlation between these two scores given that they intend to measure the same construct. A correlation of .10 was considered small, between .30 and .50 were considered medium, and .50 or above was considered large (Westen & Rosenthal, 2003). Moreover, we estimated discriminant validity to assess whether the MADRS-S measures the construct of depression and not some other construct as assessed by another instrument. To determine this, we correlated MADRS-S with STAI Y-2 and expected a smaller correlation between these two scores because they intend to measure different constructs (Cohen & Swerdlik, 2005; Cronbach & Meehl, 1955).

We performed a ROC curve analysis to ascertain MADRS-S capability to distinguish between depressed and non-depressed participants using the area under the curve (AUC). Regarding accuracy, values between .50 and .70 were considered low, between .70 and .90

were considered moderate and higher than .90 were considered high (Swets, 1988). Given the need for a gold standard measure to perform a ROC curve analysis, scores higher than 14 on the BDI-II were used as indicative of mild depressive symptomology as proposed by Beck et al., (1996).

## Results

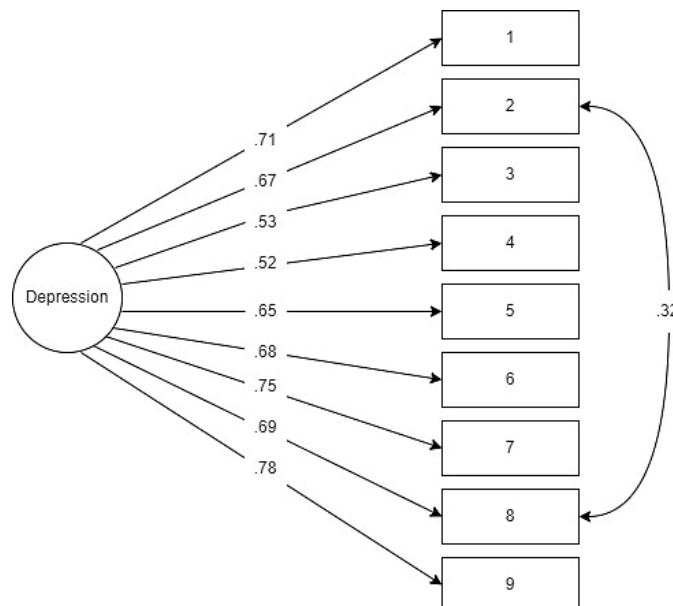
### *Confirmatory Factor Analysis*

Because data was not multivariate normal (as demonstrated by kurtosis and skewness values of  $\chi^2(143.15) = 22.41, p < .001$  and  $\chi^2(26.11) = 887.85, p < .001$ , respectively) and according to Mardia's Multivariate Normality Test, we used the maximum likelihood method with robust standard errors (MLR).

We performed CFA for a unidimensional model of MADRS-S, assuming all items loaded onto one general factor representing depression. According to the fit index thresholds, the one-factor model did not reveal an adequate fit to the data ( $\chi^2(27) = 50.70, p = .004$ , RMSEA = .07, 90% CI [.04, .09], CFI = .95). Similarly to what was considered in previous studies (Moreira & Maia, 2021), given the unidimensionality of the model and as suggested by the modification indexes, we proceeded to allow the covariation of the measurement errors of item 2 ("Feelings of unease") and item 8 ("Pessimism"). This resulted in an adequate fit to the data according to the fit index thresholds ( $\chi^2(26) = 38.35, p = .055$ , RMSEA = .05, 90% CI [.00, .08], CFI = .97). All factor loadings were above .50 as recommended by Hair et al., (2013), and statistically significant ( $p < .001$ ) ranging from .52 (item 4) to .78 (item 9).

### **Figure 1**

*Confirmatory factor analysis for the one-factor model of the European Portuguese version of the Montgomery-Asberg Depression Rating Scale Self-report (MADRS-S).  $N = 204, \chi^2(26) = 38.35, p = .055$ , comparative fit index (CFI) = .97, root mean square error of approximation (RMSEA) = .05, 90% confidence interval [.00, .08]*



### **Measurement Invariance Across Gender**

We tested the well-adjusted one-factor model for men and women separately through CFA to assess whether it fitted the data for each group. According to the fit index thresholds, the one-factor model was a good fit for women, but not for men (see **Table 2**). We proceeded to test a baseline unconstrained configural model for both groups at the same time and the results revealed a mediocre fit of the model with  $\chi^2(52) = 89.64, p < .001, RMSEA = .08, 90\% CI [.05, .11], CFI = .94$ , further indicating that the unidimensional factor structure established in the CFA did not fit the data well across men and women (configural invariance). Given that we could not determine configural invariance, we did not estimate metric or scalar invariance.

Subsequent Mann-Whitney tests showed that for men, scores on item 2 ( $Mdn = 1.00$ ) significantly differed from that for women ( $Mdn = 2.00; U = 1823.50, z = -2.68, p = .007$ ). Similarly for men, scores on item 6 ( $Mdn = 2.00$ ) significantly differed from that for women ( $M = 1.00; U = 1888.00, z = -2.46, p = .014$ ).

**Table 2**

*The goodness of fit statistics for measurement models and measurement invariance across gender for the European Portuguese version of the Montgomery-Asberg Depression Rating Scale Self-report (MADRS-S)*

	$\chi^2$	<i>df</i>	<i>p</i>	RMSEA (90% CI)	CFI
Model	50.70	27	.004	.07 [.04, .09]	.95
Modified Model	38.35	26	.055	.05 [.00, .08]	.97
Modified Model (women)	33.62	26	.145	0.04 [.00, .08]	.98
Modified Model (men)	63.21	26	< .001	0.22 [.15, .29]	.77

*Note.* RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index

### **Reliability**

We verified good internal consistency reliability as indicated by Cronbach's alpha (.88), further demonstrated by the CR value (.88). Even so, the results provided by the AVE were below the recommended value of .50 (.45). The ICC for the global score was .95 indicating a high test-retest reliability.

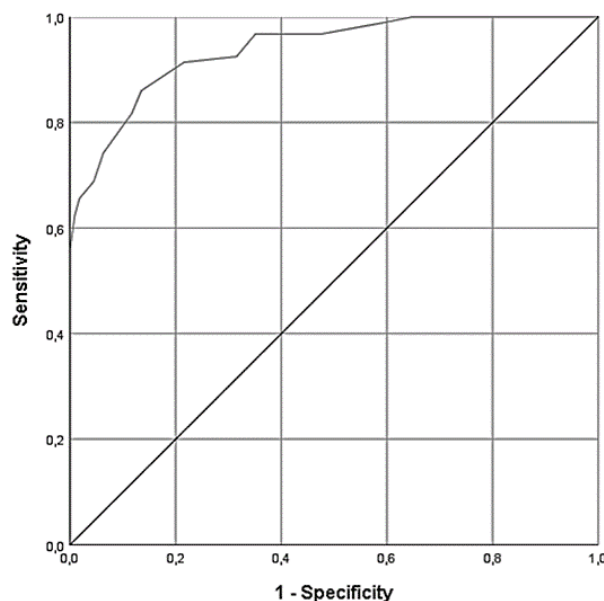
## Validity

We did not detect floor effects as only a small percentage of participants (3.4%) obtained 0 points on the MADRS-S which corresponds to the lowest possible score on the scale. Similarly, we did not observe ceiling effects given that the highest score obtained in the present study was 34 points and only one participant (0.5%) achieved that score. As expected, we were able to confirm concurrent validity between MADRS-S and BDI-II ( $r(202) = .88, p < .001$ ). We also assessed discriminant validity using the STAI Y-2 ( $r(202) = .83, p < .001$ ) where a high correlation with MADRS-S was found, suggesting that the construct measured by the MADRS-S cannot be differentiated from the construct measured by the STAI Y-2.

Criterion validity was established through a ROC analysis to evaluate MADRS-S precision in differentiating between depressed and non-depressed participants. Using the total score on MADRS-S, the AUC was calculated for the total sample revealing a high discriminatory capacity (.94, 95% CI [.91, .97],  $p < .001$ ) (Swets, 1988) (see **Figure 2**). We verified that a cut-off score of 11 allows for good sensitivity (86.0) and good specificity (86.5). Concerning gender, the AUC for women was .94 (95% CI [.90, .97],  $p < .001$ ) and .93 (95% CI [.83, 1.00],  $p < .001$ ) for men. A cut-off score of 11 shows good sensitivity (86.1) and good specificity (85.1) for women. Similarly, a cut-off score of 11 shows good sensitivity (84.6) and good specificity (94.1) for men.

## Figure 2

*Receiver operating characteristic (ROC) curve for the European Portuguese version of the Montgomery-Asberg Depression Rating Scale Self-report (MADRS-S).*



## Discussion

---

The present study describes the translation process and assesses the validity of the European Portuguese version of MADRS-S by studying its psychometric properties along with its factorial structure and diagnostic accuracy.

In line with previous studies reporting on the factorial structure of the scale (e.g., Fantino & Moore, 2009), we found that the European Portuguese version of the MADRS-S follows a unidimensional model through which all nine items load onto the latent construct of depression. The well-adjusted unidimensional model was achieved after the covariation of the measurement errors of item 2 (“Feelings of unease”) and item 8 (“Pessimism”). While item 2 relates to a general experience of anxiety and vague fear, item 8 assesses patients’ perspectives of themselves and their future with a focus on self-criticism and guilt. Repetitive negative thinking (RNT) is highly associated with self-criticism and feelings of guilt (Leonardi et al., 2020). Moreover, RNT has been considered a transdiagnostic factor that mediates the relation between anxiety and depression (Spinhoven et al., 2019). Therefore, we consider that this factor may account for the correlation between items 2 and 8 since worry and rumination (two types of RNT) have a similar association with symptoms of depression and anxiety, such as fear and pessimism (Kalmbach et al., 2016).

Through an analysis of measurement invariance across gender, we observed that the European Portuguese version of the MADRS-S is not invariant across gender, meaning that men and women respond differently to this instrument. Given that configural invariance across gender was not established, we consider that the one-factor model is not equivalent for both men and women and, therefore, the basic organization of the model is not equal across groups (Putnick & Bornstein, 2016). In an adolescent sample, Ntini et al., (2020) have previously reported gender differences on the MADRS-S regarding optimal cut-off scores. However, to our knowledge, this study presents the first gender invariance analysis for adult samples.

It is well known that men and women respond differently to measurements of depression. For instance, Martin et al., (2013) found that including non-typical depressive symptoms in which men achieve higher scores (e.g., aggression) when measuring depression, resulted in men and women meeting the necessary criteria for depression in equivalent proportions. Similarly, further studies on MADRS-S cross-gender performance should be considered to clarify gender specificities and reach a valid measure to be used both with men and women.

With this aim, we conducted subsequent preliminary analysis and identified item 2 and item 6 as having significantly different scores between men and women. In our sample, women scored higher on item 2 (“Feelings of unease”) in comparison to men while the opposite was found for item 6 (“Initiative”). The higher scores in women for item 2 (which seems to measure a general experience of anxiety) are in line with previous studies that have shown that women with depression when compared to men, receive more frequently a diagnosis of an anxiety disorder in their lifetime (Schuch et al., 2014). Additionally, women diagnosed with MDD are more likely to be diagnosed with comorbid generalized anxiety disorder across the life span (Picco et al., 2017). Regarding the higher scores by men in item 6 (which seems to measure difficulty in initiating activities), Langvik et al., (2016) suggested that men typically score higher in comparison to women on the depression scale of the Hospital Anxiety and Depression Scale (HADS) because it measures anhedonic depression. Anhedonia has been described as not only a diminished capacity to experience pleasure, but also a diminished capacity to pursue it (Rømer Thomsen et al., 2015). Therefore, men reporting higher scores on MADRS-S item 6 could be somewhat expected.

Regarding its reliability, the European Portuguese version of the MADRS-S shows good internal consistency with a Cronbach’s alpha value of .88 and a CR value also of .88. It also shows excellent temporal stability (.95). Nonetheless, AVE is below the recommended value of .50 (Hair et al., 2013), indicating that less than half of the shared variance of the items can be explained by the latent construct, and more than half is attributed to measurement error. This result is not in line with previous studies that have found a unidimensional model through which one factor explained a considerable portion of the variance (e.g., Bondolfi et al., 2010). Still, Fantino and Moore (2009) also demonstrated similar results to ours with one factor explaining only 45% of the total variance of the scale with otherwise good psychometric properties.

While we were able to confirm convergent validity of the European Portuguese version of the MADRS-S with BDI-II, we could not establish discriminant validity with STAI Y-2. Although STAI Y-2 was originally incorporated to measure trait anxiety as a discriminant measure, some debate has emerged in the last decade about the specific construct assessed by this scale. The discriminant validity of the STAI Y-2 has been called into question with serious doubts arising on whether it measures trait anxiety or if, instead, it is a nonspecific measure of negative affect and subsequently related to not only anxiety but also depressive disorders (Knowles & Olatunji, 2020). Negative affect can be described as a disposition or general tendency to experience negative emotions such as anger, guilt, and sadness even in the absence



of well-defined adverse external stimuli (Watson & Clark, 1984). Several studies support this conceptual positioning, demonstrating that STAI Y-2 has a stronger correlation with measures of depression than measures of anxiety (Bados et al., 2010; Balsamo et al., 2013).

The European Portuguese version of the MADRS-S reveals good discriminatory capacity, meaning it demonstrates satisfactory diagnostic accuracy in distinguishing non-depressed participants from depressed participants. With the European Portuguese version of the MADRS-S, it is expected that 86.0% of individuals scoring 11 points or above are true positives according to BDI-II mild depressive state and only 13.5% are false positives. This has important practical implications. For example, since this is a self-report scale, it is a measure that can be used as a screening tool in eHealth or in-person health care, complementing other diagnostic methods. Furthermore, the high specificity of the scale supports its use not only in clinical practice but also in experimental settings for the monitorization of changes in depressive symptomology across different treatment plans. Even so, different cut-off points should be utilized per the main goal of the assessment and the type of setting. For example, in examining optimal cut-off points for the BDI-II, von Glischinski et al., (2019) recommended that different scores should be applied for the screening of depression in healthy and psychiatric populations. In particular, it was noted that in psychiatric populations a higher cut-off point should be used. The same rationale should be considered for MADRS-S as well.

However, because the unidimensional model did not reveal a good fit for men, for now, we only recommend a preliminary cut-off point for both genders and particular caution using the European Portuguese version of the MADRS-S for men.

The current study is not without limitations. First, our sample included a small number of male participants which may have influenced the results observed in the measurement of invariance across gender. Future studies should employ a larger and more representative sample of the European Portuguese-speaking population, extending data collected from men, for a better and more thorough understanding of the possible influence of gender on this measure for depression.

Second, we did not establish discriminant validity of the scale with STAI Y-2 or any other distinctive measure. The discriminant validity of this scale should be tested in the future with an instrument that measures a distinctive construct with less overlapping processes with depression than anxiety (e.g., psychoticism). In fact, depression and anxiety are known to share several transdiagnostic processes such as rumination and automatic negative thoughts that support the endorsement of a dimensional model of depression and anxiety (Yapan et al., 2020).

Third, this study was limited to online participation which could have possibly led to the unintentional exclusion of potential participants with limited internet access. Although no psychometric properties of the MADRS-S seem to be affected by its use online in comparison to its paper-pencil version (Holländare et al., 2010), future studies could also evaluate if the same findings could be observed with the MADRS-S European Portuguese version.

Fourth, data collection was conducted amidst the COVID-19 pandemic which could have influenced the results of the study. In Portugal, Paulino et al., (2021) found that the psychological impact of the COVID-19 pandemic showed a positive correlation to depression, anxiety, and stress with higher scores on these constructs being reported by women. Nevertheless, while the COVID-19 context might have inflated depressive symptoms across participants, the psychometric properties of the European Portuguese version of the MADRS-S should not be jeopardized.

Although further studies are needed to confirm and add to the findings of the present study, the European Portuguese version of the MADRS-S may be considered a rapid and freely accessible self-rated measure of depression to strengthen the already available traditional evaluation instruments in clinical and experimental settings.

## References

---

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Asberg, M., Montgomery, S. A., Perris, C., Schalling, D., & Sedvall, G. (1978). A comprehensive psychopathological rating scale. *Acta Psychiatrica Scandinavica. Supplementum*, 271, 5–27. <https://doi.org/10.1111/j.1600-0447.1978.tb02357.x>
- Bacon, D. R., Sauer, P. L., & Young, M. (1995). Composite Reliability in Structural Equations Modeling. *Educational and Psychological Measurement*, 55(3), 394–406. <https://doi.org/10.1177/0013164495055003003>
- Bados, A., Gómez-Benito, J., & Balaguer, G. (2010). The State-Trait Anxiety Inventory, Trait Version: Does It Really Measure Anxiety? *Journal of Personality Assessment*, 92(6), 560–567. <https://doi.org/10.1080/00223891.2010.513295>
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., & Saggino, A. (2013). The State-Trait Anxiety Inventory: Shadows and Lights on its Construct Validity. *Journal of Psychopathology and Behavioral Assessment*, 35(4), 475–486. <https://doi.org/10.1007/s10862-013-9354-5>
- Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996). Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *Journal of Personality Assessment*, 67(3), 588–597. [https://doi.org/10.1207/s15327752jpa6703\\_13](https://doi.org/10.1207/s15327752jpa6703_13)
- Preliminary Validation Study and Psychometric Properties of the Montgomery-Asberg Depression Scale: European Portuguese Patient's Self-report Version

Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analyses of measurement studies: Measurement studies. *Ultrasound in Obstetrics and Gynecology*, 22(1), 85–93. <https://doi.org/10.1002/uog.122>

Bondolfi, G., Jermann, F., Rouget, B. W., Gex-Fabry, M., McQuillan, A., Dupont-Willemin, A., Aubry, J.-M., & Nguyen, C. (2010). Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: Evaluation in clinical practice. *Journal of Affective Disorders*, 121(3), 268–272. <https://doi.org/10.1016/j.jad.2009.06.037>

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230–258. <https://doi.org/10.1177/0049124192021002005>

Campos, R. C., & Gonçalves, B. (2011). The Portuguese Version of the Beck Depression Inventory-II (BDI-II): Preliminary Psychometric Data with Two Nonclinical Samples. *European Journal of Psychological Assessment*, 27(4), 258–264. <https://doi.org/10.1027/1015-5759/a000072>

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological Testing and Assessment: An Introduction to Tests and Measurement* (6th ed.). New York: McGraw-Hill.

Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. Psychometrika.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>

Cusin, C., Yang, H., Yeung, A., & Fava, M. (2010). Rating scales for depression. In *Handbook of clinical rating scales and assessment in psychiatry and mental health* (pp. 7–35). Humana Press.

Dimitrov, D. M. (2010). Testing for Factorial Invariance in the Context of Construct Validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121–149. <https://doi.org/10.1177/0748175610373459>

Fantino, B., & Moore, N. (2009). The self-reported Montgomery-Asberg Depression Rating Scale is a useful evaluative tool in Major Depressive Disorder. *BMC Psychiatry*, 9, 26. <https://doi.org/10.1186/1471-244X-9-26>

Fekri, O., Macarayan, E. R., & Klazinga, N. (2018). *Health System Performance Assessment in the WHO European Region: Which Domains and Indicators have been Used by Member States for its Measurement?* WHO Regional Office for Europe. <http://www.ncbi.nlm.nih.gov/books/NBK519096/>

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *JOURNAL OF MARKETING RESEARCH*, 12.

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet. Psychiatry*, 9(2), 137–150. [https://doi.org/10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3)

Greenhalgh, J., Gooding, K., Gibbons, E., Dalkin, S., Wright, J., Valderas, J., & Black, N. (2018). How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *Journal of Patient-Reported Outcomes*, 2, 42. <https://doi.org/10.1186/s41687-018-0061-6>

Gutiérrez-Rojas, L., Porrás-Segovia, A., Dunne, H., Andrade-González, N., & Cervilla, J. A. (2020). Prevalence and correlates of major depressive disorder: A systematic review. *Revista Brasileira De Psiquiatria (Sao Paulo, Brazil: 1999)*, 42(6), 657–672. <https://doi.org/10.1590/1516-4446-2020-0650>

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Pearson Education.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2013). *Multivariate data analysis: Pearson new international edition*. Harlow, Essex: Pearson Education Limited.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62. <https://doi.org/10.1136/jnnp.23.1.56>

Heo, M., Murphy, C. F., & Meyers, B. S. (2007). Relationship between the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale in depressed elderly: A meta-analysis. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *15*(10), 899–905. <https://doi.org/10.1097/JGP.0b013e318098614e>

Holländare, F., Andersson, G., & Engström, I. (2010). A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *Journal of Medical Internet Research*, *12*(5), e49. <https://doi.org/10.2196/jmir.1392>

International Test Commission. (2005). *International guidelines on computer-based and internet delivered testing*. Retrieved from [www.intestcom.org](http://www.intestcom.org)

Jain, R., Mahableshwarkar, A. R., Jacobsen, P. L., Chen, Y., & Thase, M. E. (2013). A randomized, double-blind, placebo-controlled 6-wk trial of the efficacy and tolerability of 5 mg vortioxetine in adults with major depressive disorder. *The International Journal of Neuropsychopharmacology*, *16*(2), 313–321. <https://doi.org/10.1017/S1461145712000727>

- Kalmbach, D. A., Pillai, V., & Ciesla, J. A. (2016). The correspondence of changes in depressive rumination and worry to weekly variations in affective symptoms: A test of the tripartite model of anxiety and depression in women. *Australian Journal of Psychology*, 68(1), 52–60. <https://doi.org/10.1111/ajpy.12090>
- Keilp, J. G., Ellis, S. P., Gorlyn, M., Burke, A. K., Oquendo, M. A., Mann, J. J., & Grunebaum, M. F. (2018). Suicidal ideation declines with improvement in the subjective symptoms of major depression. *Journal of Affective Disorders*, 227, 65–70. <https://doi.org/10.1016/j.jad.2017.09.018>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). Guilford Press.
- Knowles, K. A., & Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the State-Trait Anxiety Inventory. *Clinical Psychology Review*, 82, 101928. <https://doi.org/10.1016/j.cpr.2020.101928>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Langvik, E., Hjemdal, O., & Nordahl, H. M. (2016). Personality traits, gender differences and symptoms of anhedonia: What does the Hospital Anxiety and Depression Scale (HADS) measure in nonclinical settings? *Scandinavian Journal of Psychology*, 57(2), 144–151. <https://doi.org/10.1111/sjop.12272>



Leonardi, Jessica & Leonardi, J & Fimiani, Ramona & Faccini, Filippo & Gorman, Bernard & Bush, Marshall & Gazzillo, Francesco. (2020). An empirical investigation into pathological worry and rumination. *Psychology Hub*, 37(3), 31-42. 10.13133/2724-2943/17229.

Martin, L. A., Neighbors, H. W., & Griffith, D. M. (2013). The Experience of Symptoms of Depression in Men vs Women: Analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry*, 70(10), 1100. <https://doi.org/10.1001/jamapsychiatry.2013.1985>

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry: The Journal of Mental Science*, 134, 382–389. <https://doi.org/10.1192/bjp.134.4.382>

Moreira, H., & Maia, B. R. (2021). Self-critical rumination as a mediator between attachment orientations and perceived stress: Evidence from a cross-cultural validation of the self-critical rumination scale. *Current Psychology*, 40(3), 1071–1082. <https://doi.org/10.1007/s12144-018-0025-2>

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide*. (8th ed.). Los Angeles, CA: Muthén & Muthén

Ntini, I., Vadlin, S., Olofsdotter, S., Ramklint, M., Nilsson, K. W., Engström, I., & Sonnby, K. (2020). The Montgomery and Åsberg Depression Rating Scale - self-assessment for use

in adolescents: An evaluation of psychometric and diagnostic accuracy. *Nordic Journal of Psychiatry*, 74(6), 415–422. <https://doi.org/10.1080/08039488.2020.1733077>

Paulino, M., Dumas-Diniz, R., Brissos, S., Brites, R., Alho, L., Simões, Mário. R., & Silva, Carlos. F. (2021). COVID-19 in Portugal: Exploring the immediate psychological impact on the general population. *Psychology, Health & Medicine*, 26(1), 44–55. <https://doi.org/10.1080/13548506.2020.1808236>

Picco, L., Subramaniam, M., Abdin, E., Vaingankar, J. A., & Chong, S. A. (2017). Gender differences in major depressive disorder: Findings from the Singapore Mental Health Study. *Singapore Medical Journal*, 58(11), 649–655. <https://doi.org/10.11622/smedj.2016144>

Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review: DR*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Rømer Thomsen, K., Whybrow, P. C., & Kringelbach, M. L. (2015). Reconceptualizing anhedonia: Novel perspectives on balancing the pleasure networks in the human brain. *Frontiers in Behavioral Neuroscience*, 9, 49. <https://doi.org/10.3389/fnbeh.2015.00049>

Schuch, J. J. J., Roest, A. M., Nolen, W. A., Penninx, B. W. J. H., & de Jonge, P. (2014). Gender differences in major depressive disorder: Results from the Netherlands study of Preliminary Validation Study and Psychometric Properties of the Montgomery-Asberg Depression Scale: European Portuguese Patient's Self-report Version

depression and anxiety. *Journal of Affective Disorders*, 156, 156–163.  
<https://doi.org/10.1016/j.jad.2013.12.011>

Silva, D. (2003). Inventário de Estado-Traço de Ansiedade. In M. M. Gonçalves, M. R. Simões, L. S. Almeida, & C. Machado (Eds.), *Avaliação psicológica: Instrumentos validados para a população portuguesa* (Vol. 1, pp. 45-63). Quarteto.

Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2). In *Palo Alto, CA: Consulting Psychologists Press; Vol. IV.*

Spinhoven, P., van Hemert, A. M., & Penninx, B. W. (2019). Repetitive negative thinking as a mediator in prospective cross-disorder associations between anxiety and depression disorders and their symptoms. *Journal of behavior therapy and experimental psychiatry*, 63, 6–11. <https://doi.org/10.1016/j.jbtep.2018.11.007>

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

Swets, J. A. (1988). Measuring the Accuracy of Diagnostic Systems. *Science*, 240(4857), 1285–1293. <https://doi.org/10.1126/science.3287615>

- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Uher, R., Perlis, R. H., Placentino, A., Dernovšek, M. Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., & Farmer, A. (2012). Self-report and clinician-rated measures of depression severity: Can one replace the other? *Depression and Anxiety*, *29*(12), 1043–1049. <https://doi.org/10.1002/da.21993>
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- von Glischinski, M., von Brachel, R., & Hirschfeld, G. (2019). How depressed is “depressed”? A systematic review and diagnostic meta-analysis of optimal cut points for the Beck Depression Inventory revised (BDI-II). *Quality of Life Research*, *28*(5), 1111–1118. <https://doi.org/10.1007/s11136-018-2050-x>
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, *96*(3), 465–490. <https://doi.org/10.1037/0033-2909.96.3.465>

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84(3), 608–618. <https://doi.org/10.1037/0022-3514.84.3.608>

Westas, M., Lundgren, J., Andersson, G., Mourad, G., & Johansson, P. (2022). Effects of Internet-delivered cognitive behavioural therapy adapted for patients with cardiovascular disease and depression: A long-term follow-up of a randomized controlled trial at 6 and 12 months posttreatment. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, zvab131. <https://doi.org/10.1093/eurjcn/zvab131>

Wikberg, C., Nejati, S., Larsson, M. E. H., Petersson, E.-L., Westman, J., Ariai, N., Kivi, M., Eriksson, M., Eggertsen, R., Hange, D., Baigi, A., & Björkelund, C. (2015). Comparison Between the Montgomery-Asberg Depression Rating Scale-Self and the Beck Depression Inventory II in Primary Care. *The Primary Care Companion for CNS Disorders*, 17(3). <https://doi.org/10.4088/PCC.14m01758>

Wikberg, C., Pettersson, A., Westman, J., Björkelund, C., & Petersson, E.-L. (2016). Patients' perspectives on the use of the Montgomery-Asberg depression rating scale self-assessment version in primary care. *Scandinavian Journal of Primary Health Care*, 34(4), 434–442. <https://doi.org/10.1080/02813432.2016.1248635>

Yapan, S., Türkçapar, M. H., & Boysan, M. (2020). Rumination, automatic thoughts, dysfunctional attitudes, and thought suppression as transdiagnostic factors in

depression and anxiety. *Current Psychology*. <https://doi.org/10.1007/s12144-020-01086-4>

Yee, A., Yassim, A. R., Loh, H. S., Ng, C. G., & Tan, K. A. (2015). Psychometric evaluation of the Malay version of the Montgomery- Asberg Depression Rating Scale (MADRS-BM). *BMC psychiatry*, *15*, 200. <https://doi.org/10.1186/s12888-015-0587-6>