

Development and Psychometric Properties of a Novel Depression Measure

Alberto Filgueiras¹

*Language & Working Memory Lab of the University of Western Ontario, Canada
Departamento de Psicologia da Pontifícia Universidade Católica do Rio de Janeiro,
Rio de Janeiro, Brasil*

Gabriela Hora

Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil

Ana Carolina Monneratt Fioravanti-Bastos

*Programa de Pós-Graduação em Psicologia Social da Universidade Estadual
do Rio de Janeiro, Rio de Janeiro, Brasil*

Cristina M. T. Santana

Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil

Pedro Pires

Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil

Bruno de Oliveira Galvão

Centro Universitário Augusto Motta, Rio de Janeiro, Brasil

J. Landeira-Fernandez

*Departamento de Psicologia da Pontifícia Universidade Católica do Rio de Janeiro,
Rio de Janeiro, Brasil*

Abstract

The Filgueiras Depression Inventory is proposed as a new instrument, created specifically for the Brazilian culture, for screening of Major Depressive Episodes according to the categories of the DSM-V. Two studies were conducted for this purpose. Study's 1 sample consisted of 326 undergraduate psychology students. Single words or expressions were asked to represent overall depressive symptoms. The most cited formed the first version of the Filgueiras Depression Inventory. Study 2 reported the psychometric properties of this new scale. The sample consisted of 471 volunteers recruited on the Internet and 238 volunteers undergraduate students. Factor analyses, convergent and discriminant validity and other Classical Test Theory indices revealed results consistent to expectations in the present study. Andrich's Rating Scale Modeling was used as an Item Response Theory method of analysis. The overall psychometric properties of the Filgueiras Depression Inventory were shown to be good, and this study supports the effectiveness of this scale as a new instrument that measures depressive episodes in the Brazil.

Keywords: FDI, depression, item response theory, psychometrics.

¹ Endereço para correspondência: Departamento de Psicologia, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marques de São Vicente, 225, Gávea, Rio de Janeiro, RJ, Brasil 22451-041. E-mail: albertofilgueiras@gmail.com, gabrielahora@gmail.com, ana@fioravantiana.org, cristinasantan@gmail.com, ppires85@gmail.com, galvbra@yahoo.com.br e landeira@puc-rio.br.

This research was supported by a Coordination of Improvement of Higher Education Personnel (CAPES grant, Ministry of Education, Brazil) and National Council for Scientific and Technological Development (CNPq grant, Federal Government, Brazil). We thank Prof. Cilio Rosa Ziviani for valuable contributions to the Rasch Modeling and Item Response Theory.

Desenvolvimento e Propriedades Psicométricas de uma Nova Medida de Depressão

Resumo

O Inventário Filgueiras de Depressão é proposto como um novo instrumento criado especificamente para a cultura brasileira, no intuito de rastrear episódios de depressão maior de acordo com as categorias do DSM-V. Dois estudos foram conduzidos com esse propósito. O estudo 1 teve amostra constituída por 326 alunos de graduação em psicologia. Palavras únicas ou expressões foram perguntadas para que representassem os sintomas depressivos. Os termos mais citados compuseram a primeira versão do Inventário Filgueiras de Depressão. O estudo 2 reporta as propriedades psicométricas dessa nova escala. A amostra consistiu em 471 voluntários recrutados pela *internet* e 238 alunos de graduação também voluntários. Análises fatoriais, validade concorrente e convergente, e outros índices da Teoria Clássica dos Testes revelaram resultados consistentes com as expectativas do presente trabalho. A modelagem por escala de classificação de Andrich foi usada como método de análise na Teoria de Resposta ao Item. De um modo geral, as propriedades psicométricas do Inventário Filgueiras de Depressão se mostraram bons e o presente estudo sustenta a efetividade dessa nova escala como um instrumento que mede episódios depressivos no Brasil.

Palavras-chave: FDI, depressão, teoria de resposta ao item, psicomетria.

Desarollo y Propiedades Psicométricas de una Nueva Medida de la Depresión

Resumen

El Inventario de Depresión Filgueiras fue propuesto como una nueva herramienta creada específicamente para la cultura brasileña, con el fin de realizar una criba de episodios de depresión mayor según las categorías del DSM-V. Dos estudios se llevaron a cabo para este fin. El estudio 1 contó con una muestra de 326 estudiantes de licenciatura en psicología. Se pidió a palabras o expresiones para representar a los síntomas depresivos. Los términos más frecuentemente citados componen la primera versión del Inventario de Depresión Filgueiras. Estudio 2 reporta las propiedades psicométricas de esta nueva escala. La muestra fue conformada por 471 voluntarios reclutados por la *internet* y 238 estudiantes de licenciatura también voluntarios. Análisis de factores, validez convergente y discriminante, y otros índices de la Teoría Clásica de los Tests mostraron resultados que corresponderán las expectativas del presente trabajo. Se utilizó el modelo de escala de calificación de Andrich como método de análisis de la Teoría de Respuesta al Ítem. En general, las propiedades psicométricas del Inventario de Depresión Filgueiras resultó buena y este estudio apoya la eficacia de esta nueva escala como un instrumento que mide los episodios depresivos en Brasil.

Palabras clave: FDI, depresión, teoría de respuesta al ítem, psicomетria.

Major depressive disorder is one of the oldest psychopathologies ever described. According to García et al. (2006), documents that mention depression and melancholy can be found since 2,600 BC. It is also one of the most common psychiatric disorders, often resulting in impaired functioning. Suicide rates are significant-

ly higher in depressed patients compared with patients with other mental disorders. For this reason, major depression has a sad but important position among mental disorders in general (Lee & Rhee, 2003).

The 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V)

proposes several changes to the DSM-IV category of Mood Disorders (American Psychiatric Association [APA], 1994). These changes are currently being discussed, and the DSM-V was released in mid-2013 (APA, 2013). The revision alters the classification and changes the criteria of depression. In the DSM-IV, major depression is described as a unipolar disorder from the family of mood disorders related to lower mood states. It presents a series of symptoms as the main tool for a precise diagnosis. The DSM-V presents almost the same categorization, but the bereavement aspect is now removed from the exclusion criteria.

Accordingly, the APA (1994) reports that approximately 5% of the global population has major depression and associated pathologies. The descriptive symptoms in the DSM-IV and DSM-V provide nominal data about the psychological state of a patient. Major depression itself is characterized by depressive mood combined with a lack of interest or pleasure in life. The symptoms have different aspects, ranging from strictly physiological behaviors, such as changes in sleeping and eating, to mainly social behaviors, such as isolation or aggressiveness (APA, 1994, 2013; Garcia et al., 2006). To correctly diagnose the disorder of major depression according to the DSM-V, the patient needs to present two or more major depressive episodes that are separated by an interval of at least 2 consecutive months in which the criteria are not met.

Recent studies have shown comorbidity or an association of major depression with several other psychopathologies, such as anxiety (Brady & Kendall, 1992; Kemp & Felmingham, 2008; Schoevers, Beekman, Deeg, Jonker, & van Tilburg, 2003), panic syndrome (Kessler et al., 1998; Schenberg, 2010), alcoholism (Driessen et al., 2001), and eating disorders (Bulik, 2002), among others.

To achieve an accurate diagnose, Kemp and Felmingham (2008) suggested an integrative model to study emotional disorders. Their approach is based on overlapping findings in anatomic and biological measures and comparisons between depression and anxiety disorders. The

differential diagnosis of depression with regard to other psychopathologies is difficult to reach; however, self-reported instruments based on the perception of emotional reactions are still the most ethical and non-invasive way for psychologists and psychiatrists to measure emotions.

Cultural Specificities of Major Depression

Diagnosing major depression also poses the challenge of correctly sorting symptoms that appear to be influenced by culture. Epidemiological studies (Ohayon & Hong, 2006) have shown that the prevalence of major depressive disorder differs from country to country and at different moments of time. For example, an epidemiological study of major depression by Weissman et al. (1996) found that the prevalence of the disorder during one's lifetime ranged from 1.5% in Taiwan and 2.9% in Korea to 19.0% in Beirut. Prevalence rates were also different between countries, according to Ustun and Sartorius (1995). After researching 15 countries, the highest rate was discovered in South America. The United States and Europe had a moderate rate, and the lowest rate was found in East Asia.

The influences of culture, ethnicity, and different research methods were reflected in the lower estimated prevalence rates in Eastern countries than in Western countries (Kim et al., 2011). Park and Dimigen (1995) cross-culturally compared postnatal depression in Korean and Scottish mothers and concluded that the underestimation of the prevalence of depression in Korea is correlated with the social stigma of depression. According to the study, the way individuals experience the disorder is affected by the cultural context. Asians usually express depression in a somatic way, whereas North Americans and Europeans tend to manifest affective symptoms. For example, the items of the Beck Depression Inventory (BDI) have limited ability to accurately screen symptoms of depression in Korean people, because different populations appear to present some symptom variability, as discussed above (Lee & Rhee, 2003).

One of the most popular instruments to measure depressive symptoms is the BDI (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). This instrument has good reliability and validity and is widely favored in many countries (Byrne & Campbell, 1999). The BDI is a 21-question instrument with four Likert categories of discrimination (Likert, 1932). The respondent is required to choose one of four different affirmatives about his mood in the past week, including the day when the inventory is administered. The categories are then ordinally organized. Studies provide data to assure the usage of BDI as an adequate measure of depressive mood state (Mataix-Cols, Fullana, Alonso, Menchón, & Vallejo, 2004). However, the instrument is not fully able to assess traits of melancholic or depressive personalities that likely lead to recurrent depressive episodes (Vittengl, Clark, Kraft, & Jarrett, 2005).

The BDI requires the respondent to read long sentences and choose only one of the affirmatives. Semantically, deciding one predefined sentence is different from agreeing with the affirmative or a single word. Karelitz and Budescu (2004) showed how probability-related terms can be misleading. Nicol and Pickering (1993) demonstrated the role of priming in semantically ambiguous sentences and highlighted important aspects of the person's experiences with language and understanding skills.

The BDI is undoubtedly a great instrument to assess depressive episodes (Beck et al., 1961). Alternative measures are able to provide further information, different perspectives, and culturally adequate assessments of the phenomenon of major depression. Thus, we think the importance of this work is its contribution to alternative measures beyond an already established instrument, such as the BDI.

Major Depressive Episode Assessment in Brazil

Brazil has only two normalized instruments to measure depression, according to Brazil's Federal Council of Psychology (*Conselho Federal de Psicologia* [CFP], 2012). The first edition

of the BDI was culturally adapted for the Brazilian context by Cunha (2001), and the second edition (BDI-II) was normalized by Gorenstein, Pang, Argimon, and Werlang (2010). Brazil already has an instrument for the assessment of depression, the *Escala Baptista de Depressão* (EBADEP; Baptista & Gomes, 2011). It is a 17-item scale with a structure similar to the BDI's characteristics, but it also considers aspects of the World Health Organization's *International Statistical Classification of Diseases* (ICD-10) criteria for depressive episodes.

The measures in Brazil do not seem to respect an important guideline of the International Test Commission (ITC, 2010) which is being adapted to cross-cultural particularities, since to evaluate a psychological construct—such as depression; the instrument should be clear and simple enough to a target population. Several studies are suggesting an increased importance of specific measures due to cultural idiosyncrasies. Martin, Quirino, and Mari (2007) presented data supporting the idea that women in the outskirts of the city of São Paulo show that depression is a path throughout they express their sadness with poverty and violence. The difficulties faced by those women seem to lead to depressive symptoms; and health professionals should consider those cultural aspects in treatment.

Other important issue raised by this study is that psychiatrists and other mental health professionals believe that the symptomatology of depression and its representation is understood by common sense. It helps the acceptance of treatment, however some symptoms are just not associated to depression for good reason, such as aggressiveness and hopelessness—symptoms associated to their social contexts (Martin et al., 2007).

The same phenomenon happened in other study with a different population. It seems that cultural consonances—the comparison between the highest cultural consensus and the own personal cultural values; plays a mediation role between stressful events and depression (Dressler, Balieiro, Ribeiro, & dos Santos, 2007). Apparently this happens because if a person is inserted in a culture that values different cultu-

ral beliefs, she experiences inconsonance which associates with stress. The excessive presence of stress due to cultural consonance combined with the daily stress of urban environment probably leads to depression (Balieiro, Dos Santos, Dos Santos, & Dressler, 2011).

Healthcare professionals also seem to understand psychopathologies differently in Brazil when compared to other countries. Des Courtis, Lauber, Costa, and Cattapan-Ludewig (2008) studied attitudes and beliefs towards mental health patients in either healthcare services in Brazil and Switzerland. Statistically significant differences were found regarding three aspects of their beliefs: (a) Brazilian professionals are more positive towards the mentally ill than Swiss healthcaregivers; (b) Swiss professionals tend to have less knowledge of one specific case due to social distance, focusing in a psychopharmacological approach whereas Brazilian sample are more socially close to each patient leading to a more personal approach; regardless (c) Swiss healthcaregivers seem to be more open to alternative treatment strategies and social interventions than Brazilian professionals. Thus, if even among professionals of different countries there is plenty differences, probably the pathology itself is mediated by cultural aspects and relies largely in a cultural context.

Other psychopathologies also showed different symptomatology in different countries. Fontenelle, Mendlowicz, Marques, and Versiani (2004) showed that cultural factors play a significant role regarding the content of the obsessions in obsessive-compulsive disorder (OCD). Regardless of cultural differences, the core symptoms of OCD remained the same, however the treatment strategies should be adjusted according to cultural beliefs—mainly in psychotherapy which highlights the importance of culture in psychological assessment and interventions.

Other scales have been validated and shown to have psychometric properties in the Brazilian context (e.g., the State-Trait Anxiety Inventory [STAI-Y]; Fioravanti-Bastos, Santos, Maisonnette, Cruz, & Landeira-Fernandez, 2006). However, no instrument directly and exclusively

measures major depressive episode symptoms with DSM-V descriptors and one-word items or expressions. In our opinion, one-word descriptors for depressive symptoms spare valuable time in psychological assessment and are a good asset for clinical psychologists and psychiatrists.

The Filgueiras Depression Inventory (FDI) is proposed in this article as a new instrument that will help diagnose major depressive episodes (i.e., depressive mood states). It was created according to the specificities of a Brazilian sample and exclusively designed according to the DSM-V's categories. We initially studied what Brazilian psychology students think are depressive symptoms and created the FDI based on the words or expressions that these students provided that were related to major depressive episode criteria from the DSM-V. We also assessed the validity, factorability, and reliability of the scale and the items' psychometric characteristics.

The aim of this study was to design a new instrument for major depressive episode assessment and analyze the psychometric properties of this new scale using Likert-category responses based on one's assessment of a list of depressive episode symptoms.

Study 1

The first study consisted of the design and creation of the FDI. The initial proposal is a Likert-based scale created based on what a sample of psychology students from Brazil would think about symptoms of major depression.

Method

Sample

The sample of the first study consisted of 326 undergraduate psychology students (148 men [45.4%] and 178 women [54.6%]) from three large universities in Brazil: Universidade Federal do Rio de Janeiro (UFRJ; 51 men [49.6%] and 52 women [50.4%]), Universidade Federal Fluminense (UFF; 45 men [38.1%] and 73 women [61.9%]), and Pontificia Universida-

de Católica do Rio de Janeiro (PUC-Rio; 52 men [49.5%] and 53 women [50.5%]). The mean age of the sample was 23.42 years (standard deviation [SD] = 8.33). The students attended college for a mean of 3.84 years (SD = 1.10). All of the participants were voluntarily recruited from their respective institutions and had to fulfill one requirement: they had to already have taken courses in psychopathology or a related discipline.

Data Collection

The sample used in this study was chosen based on convenience. We conducted the study in three universities within the geographic range that allowed the authors to collect the data themselves. However, we did not want to ask completely naive participants because the diversity of the material that would be produced in such a case may provide only a few words or expressions that would meet the DSM-V criteria. We first explained the purpose of the research and adopted methodology to the participants. Volunteers who stated that they had never taken courses in psychopathology were excluded. We then asked the volunteers to answer a single question: "What words or expressions do you associate with a major depressive episode according to DSM-V?" After the volunteer was instructed about the study, he completed a Statement of Consent to participate in the study and answered the question by writing all of the words or expressions he thought are related to major depressive episodes. Each participant could write as many words or expressions that he wanted.

Procedure

To create the FDI items, we used the words and expressions provided by the volunteers and classified them according to the DSM-V criteria. Terms that were not associated with the DSM-V were excluded. A panel of five specialists in psychopathologies or mental health participated of the classification of the terms. The professionals had two different backgrounds: three psychologists, and two psychiatrists. All of them had at least three years of experience of clinical treatment in depression.

Once presented the list of one-word terms was presented to the specialists, each one classified the words according to the criteria established by the DSM-V for Major Depressive Episode. Communication between them were not allowed to not influence their judgment. In order to consider a word capable of describing depressive episodes according to the DSM-V, all specialists had to agree, thus, non-unanimous judgment were excluded as well as negative unanimous decisions. To represent the majority of the sample, we considered all answers that more than 50% of the sample presented, not considering possible gender differences.

Results

The terms stated by the Brazilian sample of psychology students that were related to major depressive symptoms according to the DSM-V are depicted in Table 1. Words or expressions that were not associated with any of the established criteria were excluded.

Table 2 depicts the number of participants who cited each term (i.e., percentage and by gender). The participants were allowed to say as many terms as he wanted. We conducted a χ^2 test to determine whether any significant differences existed between the concentrations of each term in either gender. No significant difference was found. We discuss these results in the General Discussion below.

All of the words associated with depression from the participants' perspective that corresponded to more than 50% of the total sample were included as an item for the FDI. To create possible answers for the respective items, we created a six-level Likert scale for self-evaluation: 1 = "Not at all related to me"; 2 = "Very little related to me"; 3 = "Little related to me"; 4 = "Moderately related to me"; 5 = "Very much related to me"; 6 = "Totally related to me."

Every word on the list received an ordinal number based on the order of the DSM-V criteria (Table 1). Afterward, the numbers were sorted randomly to generate the final structure of the instrument. Additionally, one instruction

sentence was inserted at the beginning of each item to guide the respondent through the testing experience (e.g., “Choose the option that best matches your mood state in the last fortnight”). The fortnight criterion was indicated by the differential diagnosis of the DSM-V with regard to

a major depressive episode. The items and instructions were gathered on a single paper sheet and diagrammed to be as clear as possible. The final version of this instrument was called the FDI, and a psychometric property study (Study 2) was conducted to analyze its statistical characteristics.

Table 1

Major Depressive Episode Symptoms and Respective One-Word Terms or Expressions Provided by the Sample in Study 1 (total of 36 terms)

Major Depressive Episode	One-word terms or expressions cited by study's 1 sample to compose FDI related to the DSM-V* criteria
1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, or hopeless) or observation made by others (e.g., appears tearful). Note: In children and adolescents, can be irritable mood.	Sadness, Melancholy, Bitterness, Depression, Tearfulness, Emptiness, Hopeless, Loneliness, Hate, Aggressiveness
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation)	Displeasure, Disgust, Despair, Hopelessness
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. Note: In children, consider failure to make expected weight gain	Excessive hunger, Loss of appetite, Weight loss
4. Insomnia or hypersomnia nearly every day	Bad Sleeping, Insomnia, Too much sleeping
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)	Delay, Slowness
6. Fatigue or loss of energy nearly every day	Dejection, Tiredness, Fatigue
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)	Guilt, Failure, Disappointment, Unaccomplishment, Useless
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)	Bad Thoughts, Unpleasant Thoughts, Loss of attention, Bad ideas
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide	Death, Loss of the will to live, Suicide

Note. * The DSM-V is currently under development. According to APA, publication occurs in May 2013.

Table 2

Words or Expressions Cited by the Participants from Study 1: Number and Percentage of Times Cited by Gender and Total. χ^2 Comparison between Gender to Analyze Whether the Distribution of Responses does not Present Gender Bias

Word or Expression	Number of times cited by (% of the sample)						<i>p</i> -value of the χ^2 Comparison Male x Female
	Total (<i>N</i> =326)		Male (<i>N</i> =148)		Female (<i>N</i> =178)		
Sadness	308	94.5%	138	93.2%	170	95.5%	.99
Depression	302	92.6%	131	88.5%	171	96.1%	.96
Bad thoughts	298	91.4%	123	83.1%	175	98.3%	.91
Melancholy	289	88.7%	135	91.2%	154	86.5%	.97
Disgust	278	85.3%	140	94.6%	138	77.5%	.90
Delay	264	81.0%	122	82.4%	142	79.8%	.98
Failure	253	77.6%	109	73.6%	144	80.9%	.95
Useless	241	73.9%	103	69.6%	138	77.5%	.95
Unpleasant thoughts	241	73.9%	101	68.2%	140	78.7%	.93
Slowness	235	72.1%	105	70.9%	130	73.0%	.99
Displeasure	232	71.2%	83	56.1%	149	83.7%	.82
Hate	229	70.2%	92	62.2%	137	77.0%	.90
Loneliness	224	68.7%	96	64.9%	128	71.9%	.95
Bitterness	196	60.1%	84	56.8%	112	62.9%	.96
Despair	185	56.7%	79	53.4%	106	59.6%	.95
Dejection	184	56.4%	77	52.0%	107	60.1%	.94
Guilt	181	55.5%	91	61.5%	90	50.6%	.92
Disappointment	176	54.0%	78	52.7%	98	55.1%	.98
Unaccomplishment	171	52.5%	67	45.3%	104	58.4%	.90
Death	165	50.6%	61	41.2%	104	58.4%	.86
Tearfulness	162	49.7%	63	42.6%	99	55.6%	.89
Hopelessness	161	49.4%	73	49.3%	88	49.4%	.99
Aggressiveness	160	49.1%	65	43.9%	95	53.4%	.99
Tiredness	141	43.3%	53	35.8%	88	49.4%	.88
Fatigue	136	41.7%	51	34.5%	85	47.8%	.88
Bad ideas	131	40.2%	43	29.1%	88	49.4%	.82
Loss of the will do live	93	28.5%	45	30.4%	48	27.0%	.96
Loss of attention	91	27.9%	41	27.7%	50	28.1%	1.00
Emptiness	82	25.2%	43	29.1%	39	21.9%	.92
Loss of appetite	80	24.5%	46	31.1%	34	19.1%	.87
Weight loss	73	22.4%	39	26.4%	34	19.1%	.91
Excessive hunger	54	16.6%	27	18.2%	27	15.2%	.96
Too much sleeping	52	16.0%	21	14.2%	31	17.4%	.95
Hopeless	42	12.9%	18	12.2%	24	13.5%	.98
Insomnia	40	12.3%	16	10.8%	24	13.5%	.96
Bad sleeping	26	8.0%	11	7.4%	15	8.4%	.98

Study 2

The purpose of Study 2 was to analyze the psychometric properties of the FDI based on Study 1.

Method

Sample

Two different approaches were used to collect data. A sample recruited from social networks, such as Facebook and Orkut, answered an Internet-based FDI. Another sample that consisted of undergraduate students in courses other than psychology from four universities in Rio de Janeiro and one university in Niterói answered a paper-based FDI.

The Internet-based sample had a total of 471 individuals from different regions of Brazil who voluntarily participated in the study. Of these 471 participants, 278 (59%) were female, and 193 (41%) were male. The mean age of the participants was 28.96 years ($SD = 8.94$ years), with a mean of 18.42 years ($SD = 4.40$ years) of education, equivalent to a college or university sample. Marital status was also collected. Of the 471 participants, 245 (52.0%) were single, 81 (17.2%) were currently dating or engaged, 85 (18.0%) were married, 31 (6.6%) were divorced, and 15 (3.2%) were widows/widowers.

All of the participants read the details of the study and agreed to participate after being informed of the objectives. A Statement of Consent was presented before beginning the test and signed online. This study respected the volunteers and maintained all ethical procedures.

The paper-based sample was recruited on the campus of five universities in the state of Rio de Janeiro, Brazil: UFRJ, UFF, Universidade do Estado do Rio de Janeiro (UniRIO), Universidade Estadual do Rio de Janeiro (UERJ), and Centro Universitário Augusto Motta (UNISUAM). The participants were approached in the common areas of these universities, and the researchers asked if they wanted to participate in the study. All of the participants were informed of the procedures and objectives and

signed a Statement of Consent before beginning the test.

The last sample consisted of 238 volunteers from different courses. None of the undergraduates were currently taking courses in or studying psychology. This procedure was adopted in Study 2 to avoid duplicate participation from Study 1. Of the 238 participants, 141 (59.2%) were female, and 97 (40.8%) were male. Thirty-nine were from UFRJ (16.4%), 45 were from UniRIO (18.9%), 65 were from UFF (27.3%), 41 were from UERJ (17.2%), and 48 were from UNISUAM (20.2%). For females, 25 were from UFRJ (17.7%), 31 were from UniRIO (22.0%), 33 were from UFF (23.4%), 25 were from UERJ (17.7%), and 27 were from UNISUAM (19.2%). For males, 14 were from UFRJ (14.4%), 14 were from UniRIO (14.4%), 32 were from UFF (33.0%), 16 were from UERJ (16.5%), and 21 were from UNISUAM (21.7%).

This sample was also separated by undergraduate courses. Of the 238 participants in the paper-based study of the FDI, 29 were studying engineering (12.2%), 51 were studying social communications (21.4%), 26 were studying physics (10.9%), 58 were studying economics (24.4%), 29 were studying education (12.2%), 18 were studying social services (7.6%), and 27 were studying law (11.3%).

Internet-Based Data Collection

According to Gosling, Vazire, Srivastava and John (2004), new opportunities for researchers are being provided by the fast growth of the Internet. Today, participants do not need to be in a specific room to complete self-reported scales or surveys.

A literature search was conducted that found 510 articles that used Internet-based samples with a total of 361,703 participants. The diversity of gender, social status, and other demographic data was similar to other studies without Internet-based samples. This study also presented high correlations between the results of those studies and conventional samples (Gosling et al., 2004).

Based on Gosling et al. (2004), the present study used an Internet page hosted by Google

Docs (<http://pqdepre.blogspot.com>) to collect data. We used the profiles of the researchers from two social networks (Facebook and Orkut) to recruit the participants. A Statement of Consent was first presented to the participants, and, once accepted, the study automatically began.

The FDI was posted online, together with the Brazilian Portuguese version of the BDI (Cunha, 2001), Beck Anxiety Inventory (BAI; Cunha, 2001), and STAI-Y (Fioravanti-Bastos et al., 2006). The four instruments were hosted together. A socio-demographic questionnaire was presented afterward. The final part of the website required the respondents to provide their e-mail address so that they could be contacted for feedback and the possibility to participate in further studies. One month after the first application, the participants who showed interest in continuing the study completed the FDI again.

Measures

Filgueiras Depression Inventory. The FDI is a 20-item self-report questionnaire created in Brazilian Portuguese to measure major depressive episodes. Study 1 of the present paper was conducted to create the instrument used as a measure in Study 2.

Beck Depression Inventory. The BDI is a 21-item self-report scale that assesses states of depressive mood. It is the most widely used instrument in the world for measuring depression (Beck et al., 1961; Vittengl et al., 2005). The Brazilian version of the first edition of the BDI from Cunha (2001) was used in this study.

Beck Anxiety Inventory. The BAI is a 21-item self-report inventory that assesses states of anxious mood (Beck, Epstein, Brown, & Steer, 1988). We used the Brazilian version of BAI validated by Cunha (2001).

State-Trait Anxiety Inventory-Y. The STAI-Y is a 40-item questionnaire that uses 20 items to assess states of anxious mood and another 20 items to measure traits of personality related to anxious thoughts or attitudes (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The Brazilian version of the STAI-Y was used (Fioravanti-Bastos et al., 2006).

Statistical Analysis

The FDI measures were analyzed using two different statistical approaches: Classical Test Theory (CTT) and Item Response Theory (IRT). These perspectives provide further information about the overall score and probability of response for specific items. The characteristics of the scale were considered in three aspects: (a) as a whole, (b) items one-by-one, (c) the probability of response based on the parameters from both items and participants.

Normative data were also important to understand the sample distribution based on the latent trait of depression. The mean, median, SD, 5th and 95th percentiles, skewness, and kurtosis distribution were the coefficients used to provide normative data for the FDI (Cohen, Swerdlik, & Smith, 1992). The statistical analyses were conducted using two different software programs. Descriptive statistics, and CTT indices were calculated using SPSS version 18.0 software (International Business Machines [IBM], 2009). With regard to IRT, unidimensionality test and Rating Scale Modeling were conducted using Winsteps 3.72.3 (Linacre, 2009a, 2009b).

Dimensionality

Dimensionality was assessed using variance as a Principal Component Analysis (PCA) parameter according to Rasch Family Modeling, specifically the Rating Scale Model, using Winsteps software. The PCA is not a EFA because it uses the total variance against the common variance in ML and other EFA. Regardless, Linacre (2009a) suggests to assess unidimensionality using this method if the first component explains more than 55% of the total variance, and the first contrast is under 4%.

Classical Test Theory Analyses

The CTT coefficients of the FDI were used in this study. Three different aspects were considered: (a) discriminant and convergent validity, (b) homogeneity of the test, and (c) item investigation. Discriminant and convergent validity are classic measures to determine whether the scale really evaluates what the researcher intends.

Other measures that are also considered standardized for the psychological variable were used as parameters for the scale comparison (Hambleton, Sireci, & Zumbo, 2013).

In this study, convergent validity was assessed using the BDI, and the correlation between the BDI and FDI was expected to be $> .70$. The BAI and STAI-Y were used to provide the discriminant validity coefficient. The expected moderate correlation was between $.30$ and $.70$ because the anxiety construct consistently presents those values when correlated to depression (e.g., Fioravanti-Bastos et al., 2006; Kemp & Felmingham, 2008; Schoevers et al., 2003).

The study of the homogeneity of the test was based on its internal consistency using Cronbach's alpha (Cronbach, 1951). This coefficient uses the split-half technique to estimate the correlation between two random samples of items from a universe of items that are similar to the ones in the test. Cronbach (1951) highlighted the alpha coefficient as a good index of equivalence to demonstrate the first-factor concentration of the test. A good Cronbach's alpha value is $> .70$ (Cronbach, 1969).

The investigation of the items was analyzed using item-total correlations as suggested by Cohen et al. (1992). A good item-total correlation is $> .30$. This index indicates how much an item contributes to the overall score in the scale. If an item from the FDI is $< .30$, then it likely corresponds to a latent construct that is different from the scale in which it is located.

Item Response Theory: Rating Scale Model

The Rasch analysis is a family of item-based approaches that are considered part of IRT methodology. Models from the Rasch family provide a perspective that is different from CTT and reflect latent traits that underlie an instrument. The main focus of these analyses is the usage of individual-centered statistics, quite the opposite of most psychometric methods (Rasch, 1961).

The IRT models allow researchers to test the fit between the data and theoretical model, with the advantage of being an explicit empirical model. According to Linacre (2009a), once you

have Likert-form responses in a scale and these responses present the same structure, then the Rating Scale Model is recommended. Indeed, the proposition of Andrich (1978) was to solve the problem that Rasch analyses were not suitable for polytomous items. Rasch's algorithm is adequate for dichotomous items (Rasch, 1961), but only after Andrich's Rating Scale Model were scales with responses structured as polytomous Likert items able to be analyzed using this method (Linacre, 2009a).

The main measure of the RS model is the *logit* or score location represented by the Greek letter *theta* (θ ; Linacre, 2009a). Each item presents an index of symptom severity based on the probability of an item being responded in each of the Likert categories. These values are determined by how much information about the latent trait the item is able to provide.

If an item appears to be agreed by a large amount of the sample, then it will not reveal much of a latent trait. Therefore, a large number of individuals will likely answer it as they present this symptom. In this case, the location of the item in the scale will be low. Lower *logit* values mean that is more likely to a person to agree with those items, and higher location values mean less likely to happen the same phenomenon. Good *logit* values can range from -2.0 to 2.0 . A well-balanced item presents a location value near $.0$ (Linacre, 2009a). A good scale, however, must present all types of items (i.e., both revealing and non-revealing) to successfully discriminate subjects.

The measure of fit in the RS model is given by two different misfit statistics: infit and outfit. Infit determines the inliers, which correspond to an index of adjusted residual based on the amount of symptoms presented by a person, being influenced by answers to items that present *logit* near to a person's *theta*. Outfit is also an adjusted residual index, however it is influenced by answers of a person to an item of *logit* that is far from that person's *theta*. The infit and outfit measures are given by the mean square (MnSq) fit statistics for each item and person separately, ranging from $.5$ to 1.5 to be acceptable (Linacre, 2009a).

Reliability statistics are the reproducibility of the *logit*. It is the probability that a person or item will actually have the relative measure location extracted from the relation between the empirical data and theoretical model. High person reliability shows an important amount of the RS modeling characteristics: a wide range of variance of the ability measured, long test length, high number of categories per item, and better sample-item targeting. Person reliability shows many levels of the sample the test discriminates: $> .9$ means three or four levels of discrimination, $.8-.9$ means two or three levels of discrimination, $.5-.8$ means one or two levels of discrimination, and $< .5$ means one or no levels of discrimination (Linacre, 2009a). The person reliability index can provide further information about how many percentile intervals should be used in the normative data.

Item reliability shows characteristics of the stimulus parameters. Despite being independent from test length or model fit, item reliability

varies according to two items' properties: item difficulty variance and sample size. A wider difficulty range and larger samples mean higher item reliability. However, smaller samples not only reflect lower item reliability but also cause problems when attempting to precisely locate the items on the latent variable (Linacre, 2009a). Item reliability statistics $< .8$ were considered low for the purpose of this study.

Results

Descriptive and Normative Data

Table 3 shows the mean, SD, median, 5th and 95th percentiles (scores and percentage), estimates, and standard errors of skewness and kurtosis from the FDI, BDI, BAI, and STAI-Y in each sample (i.e., Internet and paper-based). The highest score on the FDI was 120, and the lowest score was 12. This means that a lower score obtained by the subject is associated with a lower probability of depressive symptoms.

Table 3

Mean, Median, Standard Deviation (SD), 5th and 95th Percentiles (scores and percentages), and Estimates and Standard Errors (S.E.) of Kurtosis and Skewness from the FDI, BAI, BDI, and STAI-Y in Both Samples: Internet- and Paper-Based

Measure	Sample	N	Mean	Median	SD	5th Percentile		95th Percentile		Kurtosis		Skewness	
						Score	%	Score	%	Estimate	S.E.	Estimate	S.E.
FDI	Internet-based	471	53.9	51.0	23.6	29.0	15.5	109.0	4.6	1.03	.23	.38	.11
	Paper-based	238	50.4	50.0	24.9	29.0	13.3	108.0	4.1	1.08	.27	.43	.16
BDI	Internet-based	471	33.1	29.0	12.3	20.0	12.3	54.0	12.3	3.55	.23	1.74	.11
	Paper-based	238	32.0	29.0	14.5	20.0	13.1	55.0	8.6	3.56	.27	1.71	.16
BAI	Internet-based	471	23.9	21.0	18.4	3.0	14.4	59.0	9.9	-1.23	.23	-.17	.11
	Paper-based	238	21.8	20.0	17.4	4.0	15.9	56.0	10.6	-1.09	.27	-.03	.16
STAI-Y	Internet-based	471	44.5	43.0	19.3	20.0	18.1	79.0	5.1	-1.32	.23	-.20	.11
	Paper-based	238	40.9	43.0	23.8	20.0	18.9	79.0	6.7	-1.10	.27	-.11	.16

Dimensionality

In the Internet-based sample, the PCA extraction method (Rating Scale Model) presented three dimensions with eigenvalues > 1.0 . However, the variance explained by the first factor was

60.9%, which is higher than 60%. The unexplained variance of the first contrast was 3.4%, which is under 4.0%. Thus, both aspects of variance explanation indicated the FDI's unidimensionality (Linacre, 2009a). The same method in the paper-based sample also showed three dimensions

with eigenvalues > 1.0 . The variance explained by the first factor was 63.3%, and the variance of the first contrast was 2.9%. This supports the hypothesis of the possible unidimensionality of the FDI.

Classical Test Theory analyses

In the Internet-based sample, the FDI's homogeneity was measured using Cronbach's alpha and showed good internal consistency ($\alpha = .97$). Item-to-total correlations ranged from .62

for "Bad thoughts" and "Delay" to .85 for "Failure", whereas Cohen et al. (1992) suggested values $> .30$ are acceptable in a scale.

In the paper-based sample, Cronbach's alpha was slightly lower than in the Internet-based sample, but it did not diminish internal consistency ($\alpha = .94$). Item-to-total correlations ranged from .62 for "Guilt" to .81 for "Bad thoughts". Table 4 above presents means, standard deviations, item-to-total correlations, and Cronbach's alpha values if the item is deleted for all of the FDI's 20 items in both samples.

Table 4
Mean and Standard Deviation of Responses from the Internet- and Paper-Based Sample for Each Item. Item-To-Total Correlation and Cronbach's Alpha if Item is Deleted for all Items

Item	Mean		SD		Item-to-Total Correlation		Cronbach's Alpha if item is deleted	
	Internet	Paper	Internet	Paper	Internet	Paper	Internet	Paper
1. Dejection	3.59	3.70	1.57	1.62	.83	.64	.94	.91
2. Sadness	3.40	3.50	1.37	1.41	.81	.80	.95	.91
3. Failure	2.80	2.88	1.60	1.65	.85	.69	.94	.93
4. Displeasure	3.05	3.14	1.45	1.49	.78	.76	.96	.93
5. Guilt	2.77	2.85	1.49	1.53	.69	.62	.96	.91
6. Disappointment	3.34	3.44	1.66	1.71	.65	.64	.96	.91
7. Unaccomplishment	2.73	2.81	1.45	1.49	.79	.67	.96	.93
8. Melancholy	3.10	3.19	1.83	1.88	.69	.75	.96	.91
9. Bitterness	2.47	2.54	1.54	1.59	.80	.73	.95	.92
10. Disgust	2.53	2.61	1.60	1.65	.77	.64	.96	.93
11. Despair	2.15	2.21	1.45	1.41	.81	.76	.95	.91
12. Hate	1.95	2.01	1.42	1.38	.75	.74	.96	.93
13. Depression	2.56	2.64	1.55	1.51	.79	.78	.96	.93
14. Loneliness	3.05	2.99	1.58	1.54	.78	.67	.96	.91
15. Useless	2.50	2.45	1.79	1.75	.83	.74	.94	.93
16. Bad thoughts	3.07	3.01	1.45	1.41	.62	.81	.96	.91
17. Unpleasant thoughts	3.11	3.05	1.48	1.44	.78	.79	.96	.91
18. Delay	1.72	1.69	1.28	1.25	.62	.70	.96	.91
19. Slowness	2.20	2.16	1.38	1.35	.68	.79	.96	.93
20. Death	1.79	1.75	1.42	1.38	.64	.63	.96	.92

Convergent and Discriminant Validity

The convergent correlation between the FDI and BDI was high in both samples ($r = .76$ in the Internet-based sample; $r = .75$ in the paper-based sample). Therefore, the FDI data probably suggest that this scale measures depressive mood states, but in a different fashion against BDI. Discriminant correlations between the FDI, BAI, and STAI-Y were acceptable. Between the FDI and BAI, the discriminant correlation was $r = .59$ in the Internet-based sample and $r = .61$ in the paper-based sample. Between the FDI and STAI-Y, the discriminant correlation was $r = .61$ in the Internet-based sample and $r = .65$ in the paper-based sample. With regard to this criterion, moderate correlations that ranged from .30 to .70 were achieved as expected.

Rating Scale Modeling

In the Internet-based sample, person reliability was .96 and presented good sample-item targeting, good test size and length, and adequate levels of discrimination. The global RMSR was .92. This means that the relation between the expected and observed data was acceptable. Item reliability was .99, which indicates a good sample size and adequacy of the items' latent trait measures.

In the paper-based sample, person reliability was .92, which leads to the same conclusions as the previous analysis. The RMSR was .90, and item reliability was .97, indicating the adequacy and good fit of this sample with the RS model.

Misfit statistics were considered according to two criteria: infit MnSq and outfit MnSq. In both samples, infit ranged from .50 to 1.50, showing that no items degraded the scale pattern of responses. However, in the Internet-based sample, outfit ranged from .50 for "Dejection" to 1.83 for "Delay". The items "Delay" and "Death" presented values > 1.50 (1.83 and 1.80, respectively). These items appear to be the most difficult to reveal the latent trait. Even if a person is depressed, he cannot relate his mood with any of the listed symptoms. Figure 1 shows the Item Characteristic Curve (ICC) for the worst item, "Delay," and best item, "Disgust," in the RS model.

In the paper-based sample, outfit ranged from .51 for "Dejection" to 1.79 for "Delay". The same problem presented previously by the Internet-based sample appeared in the paper-based sample with regard to the outfit of "Delay" and "Death." Figure 1 also depicts the best fitted item in the paper-based sample, "Disgust", and worst fitted item, "Delay". Table 5 presents item location, infit MnSq and standard Z-score (ZStd), outfit MnSq and ZStd, and Model Standard Error for each item in the FDI in both samples.

General Discussion

In the present study, a new scale that measures major depressive episode symptoms based on DSM-V criteria contextualized to the Brazilian culture was proposed. Different measures of major depression are important to confirm the results from other measures, better understand differences between depressive episodes and major depression disorder, and use them in short time-windows of less than 30 days. The last aspect is based on the recommendations of Cronbach (1969) and Cohen et al. (1992) to not use the same measure within 1 month.

To create items for this new scale using culturally contextualized expressions, we conducted Study 1. Essentially, we asked Brazilian undergraduate psychology students what they considered major depressive episode symptoms. This procedure allowed us to develop an instrument that is adequate to the Brazilian cultural context. Nonetheless, our approach had a few limitations. Although consisting of an undergraduate sample, the participants were not completely naive to depressive episode symptoms. We used this sample due to the need of the participant's knowledge about the depression's trajectory and the established symptoms according to the DSM and ICD. Thus, the terms used to generate the FDI's items may not be entirely culturally built but constructed under an academic perspective within the Brazilian culture. However, if we wanted the full opinion of academia to build the FDI, then why did we not use statements from specialists? The intention was to create an instrument based on the nearest opinion of what the Bra-

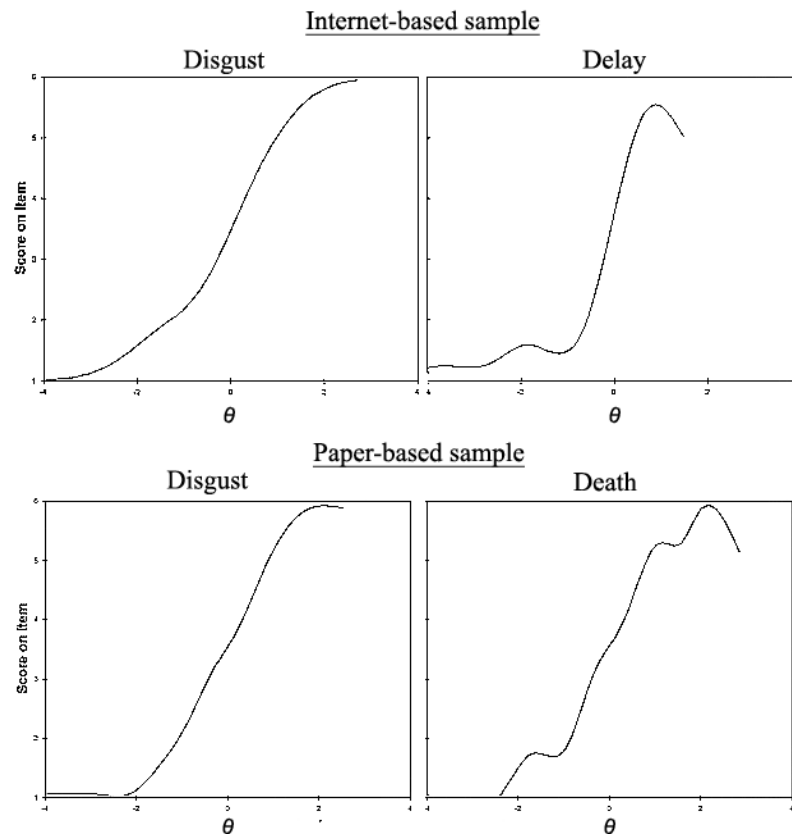


Figure 1. Item Characteristic Curve (ICC) of the best item, “Disgust”, and worst item, “Delay”, in the Rating Scale Model – in the internet-based sample; and “Disgust” – best item, and “Death” – worst item – in the paper-based sample. The criteria adopted to define the best and worst items were based on misfit coefficients. “Disgust” presented infit MnSq = 1.00 and .99; and outfit MnSq = 1.01 and .97 in the internet- and paper-based sample respectively. According to Linacre (2009a), the best value for misfit indices is 1.00, and the item “Disgust” was the closest to the desired criterion. “Delay” in the internet-based sample showed infit MnSq = 1.50 and outfit MnSq = 1.83, the farthest values from the criterion. “Death” in the paper-based sample presented infit MnSq = 1.91 and outfit MnSq = 1.76, also the worst item in misfit criterion.

zilian culture understands is a major depressive episode—something in between mental health professionals and common sense. Undergraduate students were the nearest to what we intended (i.e., not specialists). In this regard, we think the FDI achieved its goal.

To validate this new instrument, the psychometric properties of the FDI were studied using several techniques derived from CTT and IRT methods. The validity of the construct was analyzed using the convergent correlation technique with the BDI (Beck et al., 1961; Cunha, 2001) to determine whether the FDI is capable of coherently measuring Major Depression. The discriminant correlation methodology was used with the BAI (Cunha, 2001) and STAI-Y (Fioravanti-Bastos et al., 2006).

According to several studies (e.g., Beck et al., 1988; Morley, Williams, & Black, 2002), depression measured by the BDI presents two dimensions: (a) negative view of self, which involves failure, self-dislike, guilt, and punishment, and (b) somatic and physical function, including difficulty at work, loss of appetite, loss of weight, and insomnia. The DSM-V (APA, 2013) considers major depression to be mood disorder and is concerned more with low mood behaviors, such as sadness, loneliness, melancholy, bitterness, and absence of energy, than with subjective and physiological aspects, which do not load frequently from the BDI in the Factor Analysis. The FDI and BDI are instruments that use different approaches. The BDI focuses on physiological and personal visions of experien-

Table 5

Rating Scale (RS) Model Indices from Rasch Model Family in Both Samples: Internet- and Paper-Based. Location – θ , Infit, Outfit and Standard Error Statistics for Each Item. The Items are Organized in the Scale's Presentation Order

Item	Location - θ		Infit MnSq		Outfit MnSq		Model Standard Error	
	Internet	Paper	Internet	Paper	Internet	Paper	Internet	Paper
1. Dejection	-.96	-.94	.50	.53	.50	.51	.05	.05
2. Sadness	-.77	-.79	.50	.75	.71	.70	.05	.05
3. Failure	-.18	-.03	.68	.77	.73	.72	.05	.05
4. Displeasure	-.44	-.45	.69	.75	.71	.70	.05	.05
5. Guilt	-.15	-.20	1.16	1.26	1.19	1.17	.05	.05
6. Disappointment	-.72	-.73	1.40	1.49	1.41	1.38	.05	.05
7. Unaccomplishment	-.10	-.02	.77	.91	.86	.84	.05	.05
8. Melancholy	-.49	-.50	1.41	1.55	1.46	1.43	.05	.05
9. Bitterness	.20	.11	.93	.88	.83	.81	.05	.05
10. Disgust	.12	.19	1.00	.99	1.01	.97	.05	.05
11. Despair	.61	.10	.74	.89	.84	.82	.05	.05
12. Hate	.92	.95	1.26	1.14	1.08	1.06	.05	.05
13. Depression	.09	.17	.92	.95	.90	.88	.05	.05
14. Loneliness	-.44	-.45	.89	.98	.92	.90	.05	.05
15. Useless	.16	.03	1.11	1.04	.98	.96	.05	.05
16. Bad thoughts	-.45	-.46	1.27	1.51	1.42	1.39	.05	.05
17. Unpleasant thoughts	-.50	-.49	.76	.84	.79	.77	.05	.06
18. Delay	1.35	1.38	1.50	1.94	1.83	1.79	.07	.06
19. Slowness	.55	.09	1.23	1.20	1.13	1.11	.05	.06
20. Death	1.20	1.24	1.50	1.91	1.80	1.76	.06	.07

ces of depressive symptoms, and the FDI measures the low mood behaviors listed for differential diagnoses based on the DSM-V. Therefore, the Factor Analysis of the FDI is expected to present only one dimension that is strongly related to low mood behaviors.

Overall, the FDI is an unidimensional measure, and the Principal Component Analysis of residuals (PCA) in Rating Scale Model was able to provide consistent information in this regard. Some researchers proposes that samples larger than 400 participants may increase variance leading to higher number of extracted factors (Stevens, 1996). The FDI clearly has one dimension,

and other factors only appear because of the large sample and item variance. A short version of the FDI may benefit from less variance and a clearer factor analysis.

Internal consistency was measured using Cronbach's alpha (Cronbach, 1951), yielding an excellent value of $\alpha = .97$. Considering the contribution of each item to the scale, the FDI would not benefit from the deletion of any item. The item-to-total correlation ranged from .62 for "Delay" and "Bad thoughts" to .85 for "Failure". Overall, the CTT indices presented good values, and the FDI appeared to be a good and consistent measure from a psychometric perspective.

The items of the FDI were also analyzed according to the RS model under the Rasch family measurement perspective (Andrich, 1978; Linacre, 2009a; Rasch, 1961), which is an one-parameter IRT model. The FDI showed no items with infit problems and only two items with outfit problems: “Delay” and “Death”. This occurred because both items should be answered affirmatively by individuals who consistently showed responses linked to depressive symptoms. Therefore, despite being depressed, the respondents did not relate themselves to “Death” or “Delay”. This can be seen in Figure 1, which shows the lack of responses for the category “Totally related to me”.

“Delay” has two different meanings: one’s difficulties being on time and one’s slowness and reduced energy. This semantic ambiguity may be responsible for the problems found in the outfit statistics. Individuals whose scores were high related themselves only moderately with “Delay”. This likely suggests poor understanding and a problem with the item. Future studies should investigate the semantic understanding of the FDI’s items.

The DSM-V (APA, 2013) proposes a suicidal bias related to depressive symptoms. However, this aspect appears in few patients with moderate depression (Singh & Joshi, 2008). The risk of suicide in patients with depression appears to be stronger in women with high levels of depression (Allison, Roeger, Martin, & Keeves, 2001). Further studies of the FDI that use clinical samples could investigate the criteria of outfit statistics to verify whether the problem is the item “Death” or whether the study’s sample is well distributed, without a concentration of high-level depression patients, which could degrade the results of the Rasch family modeling.

The IRT analysis also allows researchers to verify person and item reliability. Person reliability is related to the number of discrimination categories, length of the scale, and levels of discrimination. The FDI’s Person reliability was .96. Therefore, the six degrees of categorical responses and 20 items were shown to be adequate for the scale. The FDI also has three or four levels of discrimination between subjects, which pro-

bably helps to correctly discriminate individuals who have depression problems from those who are at risk but do not already have depression. Future studies of sensitivity and specificity in depression patients will reveal whether the FDI accurately discriminates individuals. Item reliability is related to the adequacy of the sample size. If the empirical data respected the RS model, then the coefficient must be as close to 1.00 as possible. FDI presented item reliability = .99, which demonstrates a good sample size and sufficient variance to compare the data collected from the FDI with the logistic curve model.

Overall, the FDI’s psychometric properties were good, and the scale’s collected data showed a good ability to measure symptoms of major depression in a non-clinical sample. With regard to the items “Delay” and “Death”, both presented problems in the RS modeling but still contributed to the total score. We consider that they are good items for the FDI from a CTT perspective. The final objective of this study was accomplished because the FDI presented adequate and accurate results. However, future studies should further investigate its characteristics and possibilities.

Limitations

There is an initiative in the psychological evaluation field to universalize measures. The ITC is an example of this initiative giving guidelines for cross-cultural adaptation of instruments (ITC, 2010). FDI is a culturally based measure and should be difficult to extrapolate its results for other cultures. Other limitation is the sample of Study 1. If we intended a fully adapted assessment measure based on common sense regarding depression, then undergraduate students should not be used. Perhaps the same procedures could be used in future studies with totally naive participants from a non-specialized sample. Finally, regardless of good results, the study’s sample was non-clinical, thus, there still doubtful if the FDI is really able to identify major depressive symptoms in general population. Future studies of specificity and sensitivity should be conducted to fulfill this lack of information regarding the scale.

References

- Allison, S., Roeger, L., Martin, G., & Keeves, J. (2001). Gender differences in the relationship between depression and suicidal ideation in young adolescents. *Australian and New Zealand Journal of Psychiatry*, 35(4), 498-503. doi:10.1046/j.1440-1614.2001.00927.x.
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: American Psychiatric Press. Retrieved from <http://www.dsm5.org/Pages/Default.aspx>
- Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Balieiro, M. C., Dos Santos, M. A., Dos Santos, J. E., & Dressler, W. W. (2011). Does perceived stress mediate the effect of cultural consonance on depression? *Transcultural Psychiatry*, 48(5), 519-538. doi:10.1177/1363461511418873
- Baptista, M. N., & Gomes, J. O. (2011). Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: Evidências de validade de construto e de critério. *Psico-USF*, 16(2), 151-161. doi:10.1590/S1413-82712011000200004
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56, 893-897. doi:10.1037/0022-006X.56.6.893
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4(6), 561-571. doi:10.1001/arch-psyc.1961.01710120031004
- Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin*, 111(2), 244-255. doi:10.1037/0033-2909.111.2.244
- Bulik, C. M. (2002). Anxiety, depression, and eating disorders. In C. G. Fairburn & K. D. Brownell (Eds.), *Eating disorders and obesity: A comprehensive handbook*. New York: Guilford Press.
- Byrne, B., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-cultural Psychology*, 30(5), 555-574.
- Cohen, R. J., Swerdlik, M. E., & Smith, D. K. (1992). *Psychological testing and assessment: An introduction to tests and measurement* (2nd ed.). Mountain View, CA: Mayfield.
- Conselho Federal de Psicologia. (2012). *Sistema de Avaliação de Testes Psicológicos: Satepsi*. Recuperado em <http://www.pol.org.br/satepsi>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J. (1969). *Essentials in psychological testing* (4th ed.). New York: Harper & Row.
- Cunha, J. A. (2001). *Manual da versão em português das Escalas Beck*. São Paulo, SP: Casa do Psicólogo.
- Des Courtis, N., Lauber, C., Costa, C. T., & Cattapan-Ludewig, K. (2008). Beliefs about the mentally ill: A comparative study between healthcare professionals in Brazil and in Switzerland. *International Review of Psychiatry*, 20(6), 503-509. doi:10.1080/09540260802565125
- Dressler, W. W., Balieiro, M. C., Ribeiro, R. P., & Dos Santos, J. E. (2007). A prospective study of cultural consonance and depressive symptoms in urban Brazil. *Social Science & Medicine*, 65(10), 2058-2069. doi:10.1016/j.socscimed.2007.06.020
- Driessen, M., Meier, S., Hill, A., Wetterling, T., Lange, W., & Junghanns, K. (2001). The course of anxiety, depression and drinking behaviours after completed detoxification in alcoholics with and without comorbid anxiety and depressive disorders. *Alcohol and Alcoholism*, 36, 249-255. doi:10.1093/alcalc/36.3.249
- Fioravanti-Bastos, A. C. M., Santos, L. F., Maisonnette, S., Cruz, A. P. M., & Landeira-Fernandez, J. (2006). Avaliação da Estrutura Fatorial da Escala de Ansiedade Traço-Estado (IDATE). *Avaliação Psicológica*, 5(2), 217-224. doi:10.1007/BF02310555
- Fontenelle, L. F., Mendlowicz, M. V., Marques, C., & Versiani, M. (2004). Trans-cultural aspects of obsessive-compulsive disorder: A description of a Brazilian sample and a systematic review of international clinical studies. *Journal of Psychiatric Research*, 38(4), 403-411. doi:10.1016/j.jpsychires.2003.12.004

- Garcia, A., Passos, A., Campos, A. T., Pinheiro, E., Barroso, F., Coutinho, G., ...Sholl-Franco, A. (2006). A depressão e o processo de envelhecimento. *Science and Cognition*, 7, 111-121.
- Gorenstein, C., Pang, W. Y., Argimon, I. L., & Werlang, B. S. G. (2010). *Manual do Inventário de Depressão de Beck – BDI-II*. São Paulo, SP: Casa do Psicólogo.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, 59(2), 93-104. doi:10.1037/0003-066X.59.2.93
- Hambleton, R. K., Sireci, S. G., & Zumbo, B. D. (2013). *Psychometric methods and practices*. New York: Routledge.
- International Business Machines. (2009). IBM SPSS Amos (version 18.0). Retrieved December 27, 2012, from <http://www01.ibm.com/software/analytics/spss>; accessed
- International Test Commission. (2010). *International Test Commission Guidelines for Translating and Adapting Tests*. Retrieved from <http://www.intestcom.org/uplo ad/sitefiles/40.pdf>
- Karelitz, T. M., & Budescu, D. V. (2004). You say “probable” and I say “Likely”: improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied*, 10(1), 25-41. doi:10.1037/1076-898X.10.1.25
- Kemp, A., & Felmingham, K. (2008). The psychology and neuroscience of depression and anxiety: Towards an integrative model of emotion disorders. *Psychology & Neuroscience*, 1(2), 177-181. doi:10.3922/j.psns.2008.2.010
- Kessler, R. C., Stang, P. E., Wittchen, H. U., Ustun, T. B., Roy-Burne, P. P., & Walters, E. E. (1998). Lifetime panic-depression comorbidity in the National Comorbidity Survey. *Archives of General Psychiatry*, 55(9), 801-808. doi:10.1001/archpsyc.55.9.801
- Kim, T. S., Jeong, S. H., Kim, J. B., Lee, M. S., Kim, J. M., Yim, H. W., & Jun, T. Y. (2011). The Clinical Research Center for Depression Study: Baseline characteristics of a Korean long-term hospital-based observational and collaborative prospective cohort study. *Psychiatry Investigation*, 8, 1-8. doi:10.4306/pi.2011.8.1.1
- Lee, M. S., & Rhee, M. K. (2003). A development of Korea Depression Scale. *Journal of Korean Neuropsychiatric Association*, 42, 492-506.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140(22), 1-55.
- Linacre, J. M. A. (2009a). *User's guide WINSTEPS and MINISTEPS Rasch-model computer program*. Retrieved from www.winsteps.com
- Linacre, J. M. A. (2009b). WINSTEPS [Computer program]. Retrieved from www.winsteps.com
- Martin, D., Quirino, J., & Mari, J. (2007). Depression among women living in the outskirts of São Paulo, Southeastern Brazil. *Revista de Saúde Pública*, 41(4), 591-597. doi:10.1590/S0034-89102007000400013
- Mataix-Cols, D., Fullana, M. A., Alonso, P., Menchón, J. M., & Vallejo, J. (2004). Convergent and discriminant validity of the Yale-Brown Obsessive-Compulsive Scale Symptom Checklist. *Psychotherapy and Psychosomatics*, 73, 190-196. doi:10.1159/000076457
- Morley, S., Williams, A. C. C., & Black, S. (2002). A confirmatory factor analysis of the Beck Depression Inventory in chronic pain. *Pain*, 99, 289-298. doi:10.1016/S0304-3959(02)00137-9
- Nicol, J. L., & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research*, 22(2), 207-237. doi:10.1007/BF01067831
- Ohayon, M. M., & Hong, S. C. (2006). Prevalence of major depressive disorder in the general population of South Korea. *Journal of Psychiatric Research*, 40, 30-36. doi:10.1016/j.jpsy-chires.2005.02.003
- Park, E. H. M., & Dimigen, G. (1995). A cross-cultural comparison: Postnatal depression in Korean and Scottish mothers. *Psychologia*, 38, 199-207.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: Volume IV. Contributions to biology and problems of medicine* (pp. 321-333). Berkeley, CA: University of California Press.
- Schoevers, R. A., Beekman, A. T. F., Deeg, D. J. H., Jonker, C., & van Tilburg, W. (2003). Comorbidity and risk-patterns of depression, generalized

- anxiety disorder and mixed anxiety-depression in later life: Results from the AMSTEL study. *International Journal of Geriatric Psychiatry*, 18(11), 994-1001. doi:10.1002/gps.1001
- Schenberg, L. C. (2010). Towards a translational model of panic attack. *Psychology & Neuroscience*, 3(1), 9-37. doi:10.3922/j.psns.2010.1.003
- Singh, R., & Joshi, H. L. (2008). Suicidal ideation in relation to depression, life stress and personality among college students. *Journal of the Indian Academy of Applied Psychology*, 34(2), 259-265.
- Spielberger, C. D., Gorsuch, R., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.
- Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Ustun, T. B., & Sartorius, N. (1995). *Mental illness in general health care: An international study*. Chichester, UK: John Wiley.
- Vittengl, J. R., Clark, L. A., Kraft, D., & Jarrett, R. B. (2005). Multiple measures, methods, and moments: A factor-analytic investigation of change in depressive symptoms during acute-phase cognitive therapy for depression. *Psychological Medicine*, 35(5), 693-704. doi:10.1017/S0033291704004143
- Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H. G., ... Yeh, E. K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association*, 276, 293-299. doi:10.1001/jama.1996.03540040037030

Recebido: 27/08/2013

1ª revisão: 27/11/2013

Aceite final: 28/11/2013

Appendix

The Filgueiras Depression Inventory in Brazilian Portuguese



Nome: _____ Data de Hoje: _____

Estado Civil: _____ Idade: _____ Sexo: _____

Escolaridade: _____ Cidade/UF: _____

Este questionário contém 20 palavras que devem se relacionar com o modo como você vem se sentindo ultimamente. Uma lista de alternativas gradativas aparece primeiro, esta lista servirá para você relacionar com as palavras de conteúdo emocional logo abaixo. Leia cuidadosamente cada alternativa e palavra. Respeitando a numeração das alternativas, marque com um X o número da alternativa que melhor descreve como você vem se sentindo nos últimos 15 dias. Procure ser o mais sincero possível para retratar melhor seu estado emocional. Não há resposta certa ou errada, apenas o modo como você vem se sentindo. Não há limite de tempo, mas, em média, o preenchimento leva em torno de 10 minutos.

1. Nada relacionado comigo. 2. Pouquíssimo relacionado comigo. 3. Pouco relacionado comigo.
4. Relativamente relacionado comigo. 5. Muito relacionado comigo. 6. Totalmente relacionado comigo.

Desânimo.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Desespero.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Tristeza.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Ódio.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Fracasso.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Depressão.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Desprazer.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Solidão.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Culpa.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Inutilidade.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Decepção.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Pensamentos Ruins.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Insucesso.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Pensamentos Desagradáveis.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Melancolia.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Retardamento.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Amargura.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Lentidão.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Desgosto.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6
Morte.	<input type="checkbox"/> -1	<input type="checkbox"/> -2	<input type="checkbox"/> -3	<input type="checkbox"/> -4	<input type="checkbox"/> -5	<input type="checkbox"/> -6