

Normalization Procedure for the Baptista Depression Scale – Adult Version (EBADEP–A): Transferring of Norms

Procedimiento de normalización de la Escala Baptista de Depresión –
Versión Adulto (EBADEP–A): transferencia entre estándares
Procedimento de normalização da Escala Baptista de Depressão –
Versão Adulto (EBADEP–A): transferência entre normas

Juliana Oliveira Gomes*
Faculdade Estácio de Sá - FESJF
Makilim Nunes Baptista**
Universidade São Francisco

Doi: [dx.doi.org/10.12804/apl32.03.2014.02](https://doi.org/10.12804/apl32.03.2014.02)

Abstract

Regarding the standardization of psychological instruments, that is, the construction of referential interpretations of a test, we can find different procedures performed both by Classical Test Theory and the Theory of Item Response. Especially in this case (IRT), we can admit a test as a norm, in order to use its standardization and transfer the cut-off point to another instrument. Based on this information, the present study aimed to provide a cutoff score for the Baptista Depression Scale - Adult Version (EBADEP-A) through procedures of norms-transfer based on the Center for Epidemiologic Studies – Depression Scale (CES-D). The EBADEP-A presented good distribution and ability to discriminate depressive symptoms, and the sample, consisting of Brazilian College students, received a cutoff score of 32 points. It is emphasized that this is an exploratory and preliminary study, and it we suggest further analyzes to

be performed with clinical samples for which results can be corroborated or confronted.

Key words: Psychological assessment; test interpretation; psychometrics.

Resumen

En cuanto a la normalización de los instrumentos de evaluación psicológica, es decir, la construcción de la interpretación referencial de una prueba, podemos encontrar distintos procedimientos, hecho basado en la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI). Especialmente en este caso (TRI), podemos admitir una prueba como la norma a fin de utilizar su estandarización y transferir límites de puntuación a otro instrumento. Así, el presente estudio tuvo como objetivo proporcionar un punto de corte para la Escala Baptista de Depresión - Versión Adulto (EBADEP-A), por procedimientos de transferencia de estándares basado en la Center for Epidemiologic Studies - Depression Scale

* Juliana Oliveira Gomes, Psychology Department, Faculdade Estácio de Sá, Núcleo Zona da Mata, Juiz de Fora (MG);

** Makilim Nunes Baptista, Psychology Department, Universidade São Francisco, Itatiba (SP).

Correspondence concerning this article should be addressed to: Juliana Oliveira Gomes, Faculdade Estácio de Sá, Curso de Psicologia, Laboratório de Análises e Medidas em Psicologia (LAMP). Estrada Joaquim Vicente Guedes. Cruzeiro do Sul 36030120 - Juiz de Fora, MG - Brasil. E-mail: juogomes-psicologia@yahoo.com.br

To cite this paper: Gomes, O. J., & Baptista, N. M. (2014). Normalization procedure for the Baptista Depression Scale - Adult Version (EBADEP-A): transferring of norms. *Avances en Psicología Latinoamericana*, 32(3), 419-432. doi: [dx.doi.org/10.12804/apl32.03.2014.02](https://doi.org/10.12804/apl32.03.2014.02)

(CES-D). La EBADEP-A mostró buena distribución y capacidad de discriminar los síntomas depresivos, y la muestra de universitarios recibió un puntaje de corte igual a 32. Destacamos que se trata de un estudio exploratorio y preliminar, y se sugiere que adicionalmente se realicen análisis con muestras clínicas para que los resultados se corroboren o sean confrontados.

Palabras clave: Evaluación psicológica; interpretación de test; psicometría.

Resumo

No que concerne à normatização de instrumentos de avaliação psicológica, ou seja, a construção dos referenciais de interpretação de um teste, diferentes procedimentos podem ser encontrados, realizados com base na Teoria Clássica dos Testes e Teoria de Resposta ao Item. Especialmente neste caso, pode-se ostentar uma normatização de um teste como padrão, para transferir seus pontos de corte a outro instrumento. Assim, o presente estudo teve como objetivo apresentar uma pontuação de corte para a Escala Baptista de Depressão – Versão Adulto (EBADEP-A), por meio de procedimentos de transferência de normas com base na *Center for Epidemiologic Studies – Depression Scale* (CES-D). A EBADEP-A apresentou boa distribuição e capacidade de discriminar a sintomatologia depressiva, e recebeu a pontuação de corte igual a 32 para amostra universitária. Ressalta-se que se trata de um estudo exploratório e preliminar, e sugere-se que novas análises com amostras clínicas sejam realizadas para que os resultados sejam corroborados ou confrontados.

Palavras-chave: Avaliação psicológica; interpretação do teste; psicometria.

With respect to the construction of psychological tests, it is generally discussed the basic psychometric characteristics required for the instruments to be considered suitable for use. In Brazil, the use of psychological tests is restricted to the Psychologist, and so to ensure the minimal quality of the tests, questionnaires, scales or inventories used, the Brazilian Psychological Council, by means of a resolution (Psychologist Federal Council –CFP–, 2003), gathered the basic psychometric criteria for psychological instruments based on the Interna-

tional Test Commission (ITC, 2000) and on the Standards for Educational and Psychological Test (American Education Research Association –AERA–, American Psychology Association –APA– & National Council on Measurement in Education –NCME–, 1999). Specifically, there are two psychometric qualities that can be investigated, known as reliability and validity (Anastasi & Urbina, 2000; Noronha & Alchieri, 2004; Noronha & Vendramini, 2003; Pasquali, 1999; Urbina, 2004).

Reliability is committed to measurement error of the instrument and its stability. In other words, how much the measurement approaches to or moves away from the factual characteristics of the individual. So, the higher the accuracy of a test, greater its reliability, and lower the error in the measurement. In turn, validity regards the assumption that the test, or specifically its questions, has the ability to relate the construct which it is involved (Anastasi & Urbina, 2000; Pasquali, 1998, 1999; Urbina, 2004).

The use of a reliable and valid instrument enables the researcher, as well as the psychologist, to survey different psychological issues; even those not-so-easily observed (Noronha, 2009). This perspective of evaluation is named Classical Test Theory (CTT). The sample characteristics are influential factors to the statistical results and because of that, they're involved in reliability and validity surveys. When the researcher considers the CTT, it is possible to recover different options of reliability indices or distinct validity evidences, according to each sample investigated.

Meanwhile, considering both the historical criticism that the Psychological Assessment received over the years in Brazil, with its importance been disregarded and reconsidered for several reasons, as the ever-present need of theoretical and methodological updates, different alternative methods for the CTT procedures can be found (Amarnani, 2009; Hambleton & Jones, 1993; Pasquali & Primi, 2003). One of the theories was the Item Response Theory (IRT). The researchers suggested a rethink about both study and interpretations of a psychological instrument (Hambleton & Jones, 1993; Nunes & Primi, 2009; Pasquali & Primi, 2003; Valentini & Laros, 2011).

On the one hand, the CTT provides a methodological focus on the characteristics of the sample, since for each group investigated distinct reliability indices can be verified, and different evidences of validity, noting that reliability and validation do not decay on the test itself, but to the result interpretations of the Theory *versus* Results relationship (Urbina, 2004). On the other hand, the IRT presents a set of representation models of the test parameters, focusing both on the probability of a person to choose one or another answer according to one's skill level. This means that IRT emphasizes what may be called a Latent Trait, or the characteristic of the subject which determines the way that she or he responds to the test, in other words, the persons non-observable ability (represented by *The-ta*) (Amarnani, 2009; Embretson & Reise, 2000; Pasquali & Primi, 2003; Rueda, 2007; Valentini & Laros, 2011).

Naturally, several comparisons between CTT and IRT are expected to be found in the literature, showing their operation differences, advantages and disadvantages (Amarnani, 2009; Embretson & Reise, 2000; Fan, 1998; Hambleton & Jones, 1993; Pasquali & Primi, 2003; Rueda, 2007; Valentini & Laros, 2011; Wiberg, 2004; among others). As a result of these various comparisons, we can say that although the development of tests based on CTT cannot be considered a negative characteristic for Psychological Assessment, tests can indeed be benefited with the use of most modern and advanced statistical methods and models, as the IRT (Baptista & Gomes, 2011; Forkmann et al., 2009). Furthermore, we can also find several programs for performing the calculations, which enable to verify the residue index obtained between what was expected, and the result obtained (Nunes & Primi, 2009).

During the standardization of test procedures, i.e. the construction of the referential interpretation of the instrument (Urbina, 2004), there are also found differences related to the CTT and IRT. In the first one, the researcher needs to consider the normal distribution of the sample, and the results interpretation is connected to a main group of scores, called normative, under which the standards are constructed (Anastasi & Urbina, 2000; Urbi-

na, 2004). In the second one, in turn, assuming the focus on latent trait and on the instrument, we can admit the standardization of a test by considering one instrument as the default, and later transfer its cutting points to the other one (Thomas, 2011).

For such a procedure, items and persons present in the database are subjected to a calibration, so that the parameters are isolated, and therefore they lose their dependence on each other. Furthermore, measurements of both instruments are also calibrated on a common scale, allowing the comparison and the transfer of normative standards from one to another, through the equalization procedure (Bauer & Husong, 2009; Embretson & Reise, 2000; Thomas, 2011). The equalization process allows instruments that theoretically measure the same construct, to become comparable and proportional, making it possible that the same statistical significance (and so their interpretations) to be attributed to participants with the same ability (Smith et al., 2006; Wyse & Reckase, 2011).

Although IRT is presented with a predominant focus on skills and abilities, like in trial-and-error instruments, we might also notice studies using these models to assess more subjective constructs, related to the mental health area, such as depression for example (Castro, Trentini, & Riboldi, 2010; Cole et al., 2011; Covic, Pallant, Cnaghan, & Tennant, 2007; Covic, Pallant, Rabin, & Kaufman, 2004; Jones & Fonda, 2004; Pickard, Dalal, & Bushnell, 2006; Sauer, Ziegler, & Schmitt, 2012; among others). IRT methods can, indeed, provide many benefits on subjective or complex psychopathological measures (Reise & Waller, 2009).

About the depression measurement, several instruments can be presented, many of which with validity and reliability evidences, in different countries, including Brazil (Calil & Pires, 1998; Santor, Gregus & Welch, 2006). A widely used instrument is the Center for Epidemiologic Studies - Depression Scale (CES-D) which assesses the construct through 20 items, with an emphasis on affective components and depressed mood (Radloff, 1977). Specifically in Brazil, we can identify different surveys about validity evidences, from which distinct cutoff points were established, according to the samples investigated, that is, college students

(cutoff 15), drugs addicted groups (cutoff 16) and general population (cutoff 24) (Batistoni, Neri & Cupertino, 2007; 2010; Hauck Filho & Teixeira, 2011; Silveira & Jorge, 1998).

The use of imported and translated instruments, since they possess the necessary adaptations, reliability, and validity studies, besides different samples of reference, can benefit both the scientific community and clinicians, whom can use them in different contexts (Urbina, 2004). However, it is always essential to underline the importance of not only focus on updating and adapting imported scales and questionnaires, but also emphasize the construction of new instruments, with contextualized items intended for different populations.

In this direction, it is highlighted the Brazilian instrument named Baptista Depression Scale - Adult Version (*Escala Baptista de Depressão - Versão Adulto*; EBADEP-A), which focuses on the assessment of depressive symptoms based on 26 descriptors, over 45 items (Baptista, 2012). The EBADEP-A has evidences of validity and reliability based both in CTT and IRT, and it also presents ratings of the symptoms severity, namely minor/ none symptomatology of depression, mild, moderate and severe symptomatology, established by the transfer of norms procedure from the Beck Depression Inventory (Baptista, 2012; Baptista & Gomes, 2011; Baptista, Cardoso, & Gomes, 2012).

This study assumes that the IRT allows, more accurately than the CTT, the reduction of errors on measurements in the analysis of latent variables, once the model considers not only expected responses, but also some unexpected, represented by the INFIT (which attenuates the importance of external residues) and OUTFIT (which is more sensitive to external residues) indices. Moreover, it is assumed that it is possible to assess the bias and to analyze the adjustments of the items of these instruments, as well as use calibration and equalization of these measures, so that they might be comparable (Embretson & Reise, 2000; Thomas, 2011). Therefore, the present study presented as a research problem the possibility of a cutoff point based on this accurate measurements. Thus, this article aimed to provide a cutoff point for the Baptista Depression Scale - Adult Version, through transfer of norms

procedures from the Center for Epidemiologic Studies - Depression Scale (CES-D).

Method

Participants

The study included 589 college students from different cities in two Brazilian states, São Paulo and Minas Gerais. From this total, 519 (88%) were from undergraduate courses, such as Nursing, Geography, Psychology, Law School, Biology, Nutrition, Pharmacy and Physiotherapy, and 43 (7.4%) were from postgraduate courses in Psychology, while 27 (4.6%) did not respond. The sample was composed predominantly of women ($n = 439$, 74.5%), and ages ranged from 17 to 63 years with mode equal to 18 ($M = 23.48$, $SD = 8.56$). Regarding marital status, the majority were single ($n = 495$, 83.9%).

Instruments

Baptista Depression Scale - Adult Version (EBADEP-A). The EBADEP-A was constructed by Baptista (2012), from 26 descriptors of depressive symptoms, based on the international diagnostic manuals CID 10 and DSM-IV-TR, as well as in Cognitive and Behavioral theories of depression (APA, 2002; Beck, Rush, Shaw, & Emery, 1997; Ferster, Culbertson & Boren, 1977; OMS, 1993). It has 45 four-point Likert items, and it consists of pairs of sentences with positive and negative imprints which refer to the descriptors of depression. Its minimum score is zero and the maximum 135, and the higher its score, the greater depression symptoms presented.

The instrument has several published evidences of validity and reliability, based both in CTT and IRT (Baptista, 2012; Baptista & Gomes, 2011; Baptista, Cardoso, & Gomes, 2012). Baptista and Gomes (2011) analyzed the psychometric qualities of the instrument, in a validity study of construct and criterion evidences, based on both CTT and in IRT. As for CTT the ANOVA determinate good group discrimination, in which included the college students group, psychiatric patients, and depressi-

ve and non-depressed people. Regarding the TRI, the internal structure of the instrument was considered appropriate and the study of Differential Item Functioning showed no issues. Finally, the Cronbach's alpha ($\alpha = 0.95$) and the index generated from the Rasch model ($\alpha = 0.92$) were considered excellent, indicating good accuracy of reliability. In another essay, Baptista, Cardoso, and Gomes (2012) investigated evidence of convergent validity between this instrument and the Beck Depression Inventory (BDI-II), finding good and significant correlations in the study of temporal stability ($r = .69$; $p \leq 0.000$), with one month intermission.

Center for Epidemiologic Studies – Depression Scale (CES-D). The CES-D is a screening scale of depressive symptomatology, composed of 20 four-point Likert items. It has been built based on different instruments for depression, in order to be used both in clinical and non-clinical contexts. Its minimum score is zero and the maximum 60, and the higher the overall score, the greater presence of depressive symptoms (Radloff, 1977). In Brazil, it was translated by Silveira and Jorge (1998), who presented the following scores: minimum of 15 for the presence of depressive symptoms in college students, 16 for substance abusers and 24 for the general population.

In Brazil, different surveys using this instrument can be found, from which we can present evidences of validity and reliability (Batistoni, Neri & Cupertino, 2007; 2010; Hauck Filho & Teixeira, 2011; Silveira & Jorge, 1998). In a study with young population, Jorge and Silveira (1998) applied the instrument in two groups. The first one, composed of College students, composing a non-clinical sample, and the second group, clinical, composed of young people with substance use disorders. The evidences of concurrent validity, internal consistency and factor structure were verified. The instrument was divided into four factors, similar to the original test in English. Finally, Cronbach's alpha of 0.87 was determined. Regarding the study with elderly population, Batistoni, Neli and Cupertino (2007) conducted a research for evidences of internal consistency, construct and criterion. It were evidenced satisfactory levels of internal validity ($\alpha = 0.86$),

sensitivity (74.6%) and specificity (73.6%). Finally, factor analysis generated a division in three factors for this population.

Procedures

To perform this psychometric research, all ethical procedures were adopted. The project was approved by one Ethics Committee (Universidade São Francisco, Itatiba, Brazil), and before answering the questionnaires each participant signed, in two copies, a Term of Consent. An Identification Questionnaire with general information was presented to the sample, followed by the EBADEP-A, and the CES-DA. The instruments were answered in approximately 35 minutes. As inclusion criteria, we only accepted protocols that presented minimally 80% of the instruments properly answered i. e. with less than three non-responded questions.

Results

First, we checked the inclusion criteria and excluded protocols with number greater or equal to three blank answers on one of the two instruments. As regards to the EBADEP-A, the question with more missing data was “completing tasks” ($n = 13$), particularly linked to the descriptors of feelings of inadequacy and loss of productivity, followed by issues related to loss of libido ($n = 8$), helplessness, neediness, depressed mood, lack of perspective with the present ($n = 6$), and fatigue/ energy loss ($n = 4$). In turn, for the CES-D the largest number of blank answers was the lack of perspective with the future ($n = 6$), followed by helplessness ($n = 5$), irritability, and anhedonia ($n = 4$). All blank questions were replaced by individual averages.

Then, in a descriptive way, the structures of each instrument, and data adequacy to the model were verified. Regarding to items, we first established the EBADEP-A infit mean values ($M = 1.03$, $SD = 0.26$) and outfit mean values ($M = 1.02$, $SD = 0.28$), just as the CES-D infit mean values ($M = 1.00$, $SD = 0.32$), and outfit mean values ($M = 1.02$, $SD = 0.51$). This data indicated that both instruments were answered in a default line and are suitable to the Rasch model (Linacre, 2010). Specifically

about the distribution of the response options, it is presented in the Figure 1 the graphical display related to both scales.

We might see from Figure 1 that, although the scales have the same Likert options (from zero to

three), the distribution does not occur in the same path, since the thresholds between options are different. To the EBADEP-A, they ranged from -2.52 to 2.32, while for the CES-D the variation was between -1.83 and 1.83. Meanwhile, it was revealed

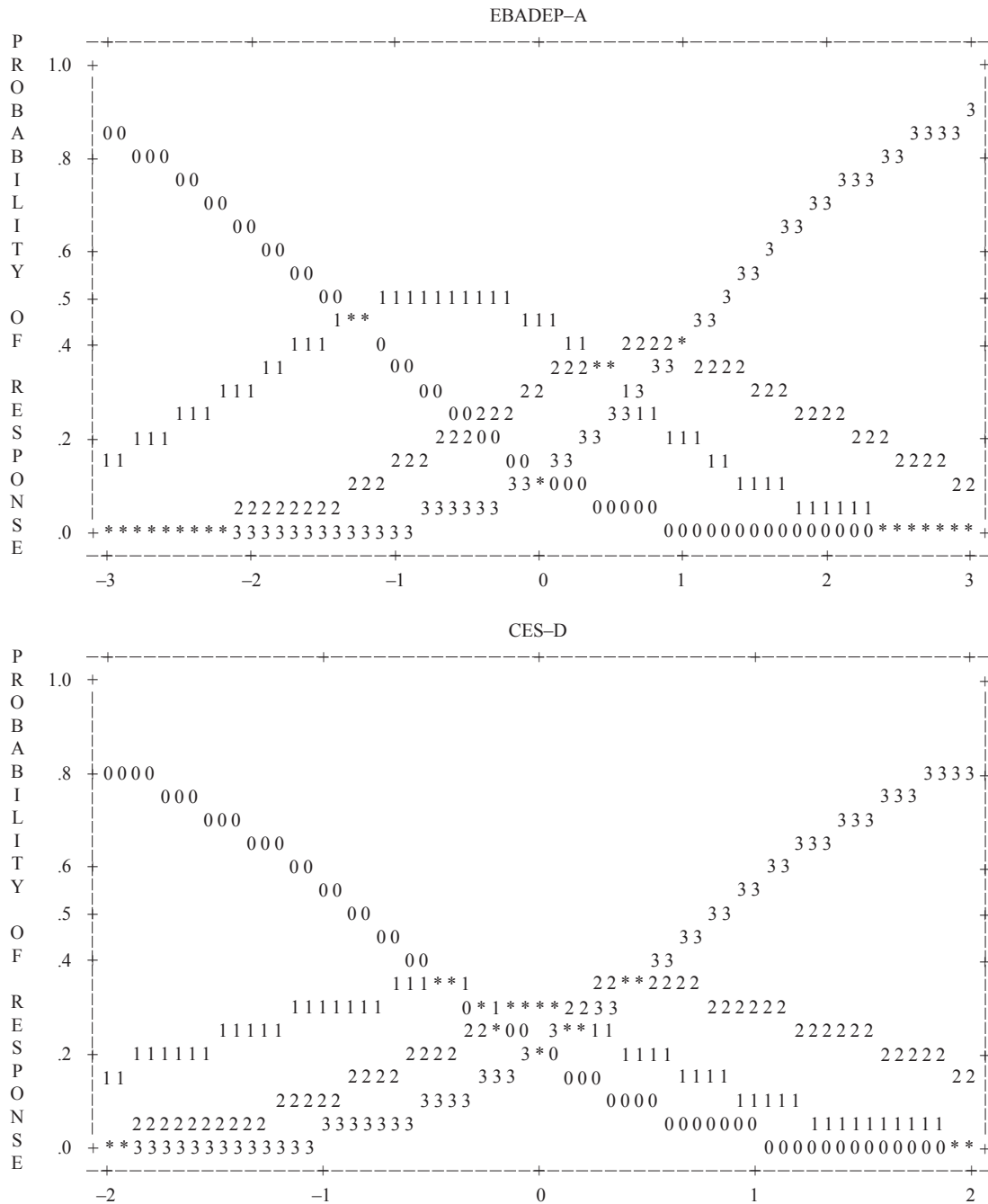


Figure 1. Distribution of response options of the instruments

that the instruments and the data showed fit to the Rasch model, what demonstrated the possibility of performing the transfer procedures.

In order to proceed, both scales were analyzed together, treated as a single set of items. Through the map of items (Table 1) it was possible to explore such structure, based on the graphical display of skill levels of persons, and with regard to the items difficulty, that is, concerning to the assessment of people which showed more or less severe depression, according to the instruments scores.

The map of items (Table 1) shows the variation and distribution of the items difficulties, as well as the ability levels of persons. The letters M located in the center column indicate the average positions. We observed that the average value of the items was larger than the average value of persons, and therefore, we can understand that the items were rated as more difficult than the abilities. Since EBADEP-A and CES-D are not ability instruments, with right-and-wrong questions, it can be said that most questions received the minimum values possible, and all the items were positioned above the average of persons.

First, the item considered more difficult was related to suicidal ideation (EBADEP29), what shows that for this item, the participants had chosen most frequently the circles next to the positive sentence in the instrument. Then, three items of the CES-D, referring to perspectives about the past, present and future (CESD16, CESD8, and CESD12), could be consider as easier, i.e., had been widely reported with the maximum values. In a general context, we concluded that although EBADEP-A has more than twice numbers of items than the CES-D, this one presented easier items, with greater possibility of maximum responses.

With the porpoise to analyze the instruments on the same scale, it was necessary to calibrate and equalize the measures of test parameters, so that the Theta values could be used to the transfer the cut-off point from the CES-D to EBADEP-A (Bauer & Hussong, 2009; Embretson & Reise, 2000; Smith et al., 2006; Thomas, 2011; Wyse & Reckase, 2011). After the procedures of equalization and anchorage, all possible CES-D scores (from zero to 60) were associated with an ability level, from which we could

Table 1
Map of Items for CES-D and EBADEP-A

Persons – Items
<More> <Rare>
2 +
EBADEP29
. T
EBADEP44
. CESD9
1 . + CESD2 EBADEP10
. EBADEP12 EBADEP33 EBADEP34
S CESD15 CESD3 EBADEP30
. CESD17 CESD19 CESD20 EBADEP21
T CESD1 EBADEP13 EBADEP28 EBADEP32
EBADEP11 EBADEP41 EBADEP42 EBADEP5 EBADEP9
CESD10 CESD6 EBADEP20 EBADEP35 EBADEP43
EBADEP8
0 ## +M CESD13 EBADEP16 EBADEP17 EBADEP19 EBADEP22
EBADEP3 EBADEP31 EBADEP4
CESD18 EBADEP18 EBADEP24 EBADEP40
S CESD11 CESD14 EBADEP39 EBADEP6
CESD7 EBADEP23 EBADEP25 EBADEP26 EBADEP7
CESD4 EBADEP2 EBADEP27 EBADEP36
S CESD5 EBADEP1 EBADEP15 EBADEP45
EBADEP14 EBADEP37 EBADEP38
-1 ##### +
M
#####
T
#####
CESD16 CESD8
#####
-2 ##### S+ CESD12
#####
.#
###
.
T
.
-3 .# +
.
.
.
-4 +
.
-5 . +
<less> <freq>

Each “#” is 4. Each “.” is 1 to 3

verify the equivalent expected measure to the cutoff point to this sample, of college students (Table 2).

Thereby, in table 2, it was possible to find out that the Theta value associated to the cutoff point 15, established by Silveira and Jorge (1998), is -1.17. Proceeding, in order to transfer such CES-D standard to EBADEP-A, all the possible values of this scale (from zero to 135), were also associated with levels of ability (Table 3). In this case, we needed to identify an equivalency between the cutoff point 15 and the ability level presented as Theta score. So to do that, we scanned all the possible values on Table 3 and searched a EBADEP-A Theta measure that could be correspondent to the CES-D cutoff score above mentioned. For the Theta ability -1.17 it has been found the associated score of 32.

As a additional information, through crosstabulation, it was possible to evidence that 269 participants presented a of depressive symptoms by CES-D, while 222 (82.5 %) showed the same diagnosis by EBADEP-A. In Table 4, we could observe that the coincidence on depression diagnosis occurred in 65 % of all cases. It was also determinated the association between variables, using the chi-square test ($X^2 = 134.47$; $df = 1$; $p \leq 0,001$) and Pearson's correlation ($r = .71$; $p \leq 0,001$).

Finally, it is emphasized that the cutoff point of the CES-D for college students (15) equals to 25 % of the total possible scores of the instrument. After the transfer of norms, the EBADEP-A received cutoff score for college students equal to 32, i.e. 23.7% of all possible scores. Originally, 320 students (54.3 %) had depressive symptoms according to the CESD-D cut-off point. After the transfer of norms, we found that 255 college students (43.3 %) could be classified as depressed by the EBADEP-A.

Discussion

Within the context of psychological instruments construction, it is always emphasized their basic psychometric qualities, validity and reliability, so they can be used in different contexts (Anastasi & Urbina, 2000; Noronha & Alchieri, 2004; Noronha & Vendramini, 2003; Pasquali, 1999; Urbina, 2004). For both procedures and techniques, one can use statistical models based in both Classical

Test Theory (CTT) and the Item Response Theory (IRT).

The same argument applies to procedures of standardization, i.e. the interpretive rules for the test (Urbina, 2004). On one hand, the CTT is based on the characteristics of sample groups, while on the other hand, the IRT focuses on the characteristics of the items, and the probability of participants in sample to indicate one or another kind of response, with emphasis on the latent trait (Amarnani, 2009; Embretson & Reise, 2000; Pasquali & Primi, 2003; Rueda, 2007; Valentini & Laros, 2011).

Indeed, there are different studies which seek to show the advantages, disadvantages and differences between the CTT and the IRT (Amarnani, 2009; Embretson & Reise, 2000; Fan, 1998; Pasquali & Primi, 2003; Rueda, 2007; Valentini & Laros, 2011; Wiberg, 2004; among others). However, it is noteworthy that although the use of CTT in Psychological Assessment is not a downside to the evolution of the clinical and nonclinical Psychology, and for the progress of the surveys, the use of the techniques shown by models of IRT can be benefit once they use more advanced statistical methods, with a reduced number of noise and waste (Baptista & Gomes, 2011; Nunes & Primi, 2009; Forkmann et al., 2009).

One of the possible procedures for the standardization of an instrument by IRT, involves the transfer of norms of a standards previously valid and reliable test, by comparing the level of ability required to achieve the cutoff score set (Thomas, 2011). For this, instruments must be measured within the same scale, i.e equalized, and subsequently calibrated, in order to make it possible to transfer each instruments into Theta.

Under these assumptions, the aim of this study was to realize a transfer of norms from the Center for Epidemiologic Studies - Depression Scale (CES-D), a rating scale of depressive symptoms widely known and published in different countries, including Brazil (Batistoni et al., 2007; 2010; Hauck Filho & Teixeira, 2011; Radloff, 1977; Silveira & Jorge, 1998), to the Baptista Depression Scale - Adult Version (EBADEP-A), which was built in Brazil and presents different evidences of validity and reliability in the country (Baptista, 2012; Baptista & Gomes, 2011; Baptista et al., 2012).

Normalization Procedure for the Baptista Depression Scale – Adult Version (EBADEP-A): Transferring of Norms ■

Table 2
Conversion of scores into Thetas for CES-D

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
0	-5.29E	1.82	21	-0.66	0.28	42	0.76	0.27
1	-4.10	0.99	22	-0.59	0.27	43	0.84	0.27
2	-3.42	0.70	23	-0.52	0.27	44	0.91	0.28
3	-3.02	0.58	24	-0.44	0.27	45	0.99	0.28
4	-2.73	0.51	25	-0.37	0.26	46	1.08	0.29
5	-2.49	0.46	26	-0.31	0.26	47	1.16	0.30
6	-2.30	0.43	27	-0.24	0.26	48	1.25	0.31
7	-2.13	0.40	28	-0.17	0.26	49	1.35	0.32
8	-1.98	0.38	29	-0.10	0.26	50	1.45	0.33
9	-1.84	0.37	30	-0.04	0.26	51	1.57	0.34
10	-1.71	0.35	31	0.03	0.26	52	1.69	0.36
11	-1.59	0.34	32	0.09	0.25	53	1.83	0.38
12	-1.47	0.33	33	0.16	0.25	54	1.98	0.41
13	-1.37	0.32	34	0.22	0.26	55	2.17	0.45
14	-1.27	0.31	35	0.29	0.26	56	2.39	0.50
15	-1.17	0.31	36	0.35	0.26	57	2.67	0.57
16	-1.08	0.30	37	0.42	0.26	58	3.07	0.70
17	-0.99	0.29	38	0.49	0.26	59	3.76	1.00
18	-0.90	0.29	39	0.55	0.26	60	4.96E	1.82
19	-0.82	0.28	40	0.62	0.26			
20	-0.74	0.28	41	0.69	0.27			

Table 3
Conversion of scores into Thetas for EBADEP-A

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
0	-6.32E	1.83	46	-0.6	0.19	92	0.88	0.18
1	-5.10	1.01	47	-0.56	0.19	93	0.91	0.19
2	-4.39	0.72	48	-0.53	0.19	94	0.95	0.19
3	-3.97	0.59	49	-0.49	0.19	95	0.98	0.19
4	-3.67	0.52	50	-0.46	0.19	96	1.02	0.19
5	-3.43	0.46	51	-0.42	0.18	97	1.06	0.19
6	-3.23	0.43	52	-0.39	0.18	98	1.09	0.19
7	-3.06	0.40	53	-0.36	0.18	99	1.13	0.19

Continúa

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
8	-2.91	0.38	54	-0.32	0.18	100	1.17	0.19
9	-2.78	0.36	55	-0.29	0.18	101	1.20	0.20
10	-2.65	0.34	56	-0.26	0.18	102	1.24	0.20
11	-2.54	0.33	57	-0.23	0.18	103	1.28	0.20
12	-2.44	0.31	58	-0.19	0.18	104	1.32	0.20
13	-2.34	0.30	59	-0.16	0.18	105	1.36	0.20
14	-2.26	0.29	60	-0.13	0.18	106	1.41	0.21
15	-2.17	0.29	61	-0.10	0.18	107	1.45	0.21
16	-2.09	0.28	62	-0.07	0.18	108	1.49	0.21
17	-2.01	0.27	63	-0.03	0.18	109	1.54	0.21
18	-1.94	0.27	64	0.00	0.18	110	1.59	0.22
19	-1.87	0.26	65	0.03	0.18	111	1.63	0.22
20	-1.81	0.26	66	0.06	0.18	112	1.68	0.22
21	-1.74	0.25	67	0.09	0.18	113	1.74	0.23
22	-1.68	0.25	68	0.12	0.18	114	1.79	0.23
23	-1.62	0.24	69	0.15	0.18	115	1.84	0.24
24	-1.56	0.24	70	0.18	0.18	116	1.9	0.24
25	-1.51	0.23	71	0.21	0.18	117	1.96	0.25
26	-1.46	0.23	72	0.24	0.18	118	2.02	0.25
27	-1.40	0.23	73	0.28	0.18	119	2.09	0.26
28	-1.35	0.22	74	0.31	0.18	120	2.16	0.27
29	-1.30	0.22	75	0.34	0.18	121	2.23	0.28
30	-1.25	0.22	76	0.37	0.18	122	2.31	0.28
31	-1.21	0.22	77	0.40	0.18	123	2.4	0.29
32	-1.16	0.21	78	0.43	0.18	124	2.49	0.31
33	-1.12	0.21	79	0.46	0.18	125	2.58	0.32
34	-1.07	0.21	80	0.49	0.18	126	2.69	0.34
35	-1.03	0.21	81	0.52	0.18	127	2.81	0.36
36	-0.99	0.20	82	0.55	0.18	128	2.95	0.38
37	-0.94	0.20	83	0.59	0.18	129	3.10	0.41
38	-0.90	0.20	84	0.62	0.18	130	3.28	0.45
39	-0.86	0.20	85	0.65	0.18	131	3.51	0.50
40	-0.82	0.20	86	0.68	0.18	132	3.79	0.57
41	-0.79	0.20	87	0.71	0.18	133	4.19	0.70

Continúa

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
42	-0.75	0.19	88	0.75	0.18	134	4.88	1.00
43	-0.71	0.19	89	0.78	0.18	135	6.08E	1.82
44	-0.67	0.19	90	0.81	0.18			
45	-0.64	0.19	91	0.85	0.18			

Table 4
Crosstabulation between the diagnostic categories of the instruments

		CES-D		Total	
		Non depressed	Depressed		
EBADEP-A	Non depressed	N	222	112	334
		%	82.5%	35.0%	56.7%
	Depressed	N	47	208	255
		%	17.5%	65.0%	43.3%
Total	N	269	320	589	
	%	100.0%	100.0%	100.0%	

Initially, the model adequation indices were verified by the average values of infits and outfit (Linacre, 2010). It was also observed that both tests showed good distribution of their response options. Although both instruments present the same type of Likert scale, between zero and three points, the structures are, indeed, very different, and for this reason, for the transfer of cutoff points the initial procedures of equalization and anchorage were conducted.

Through the interpretation of map of items, in which both scales were evaluated together, we could see that the items were considered difficult to the sample, since most of them were located above the population average. This trait indicates that there was greater adherence of the sample for answering lower choices, near zero, suggesting low depressive symptoms of the persons, which is somewhat expected in a nonclinical population. So, we can say that the EBADEP-A is effectively evaluating depressive symptoms in college students, once if the extent of items and people were very similar, the scale would be classified as very easy for the sample, and consequently, it would not be captured adequately depressive symptoms.

A similar procedure of transfers of norms was accomplished by Baptista (2012), based on the Beck Depression Inventory (BDI). Most items of EBADEP-A were below the average of the items and above average of persons, showing that the EBADEP-A presented the ability to differentiate moderate from severe symptoms, and proving that the BDI differed better milder symptoms from the moderate. In the present research, although EBADEP-A nature has not exclusive screening, as with the CES-D, it showed a good ability to track depressive symptoms in college students.

Through data equalization, both instruments involved in the present study began to exhibit the same scale of difficulty, and therefore, the ability levels (Theta) could be interpreted as equivalent. Initially we set every possible scores of CES-D, in order to track down the equivalent Theta to the scale cut-off point. Later, the same procedure was performed for the EBADEP-A, so that from Theta established in the previous procedure, it could be endorsed the new cutoff score.

Thus, we can conclude that the EBADEP-A showed good screening feature of depressive symptoms in the sample, and it had received a cutoff

equal to 32 to the presence of depressive symptoms. However, it should be highlighted that this is an exploratory study, and therefore the results should be treated cautiously. First of all, about the CES-D, it does not provide points of severity of depressive symptoms, as in the BDI or EBADEP-A, but only one cutoff point, which may limit the interpretations. Furthermore, it raises questions about the sample type used, i.e. participants whose clinical diagnosis of depression was not measured: can the sample selection, with a low frequency of symptoms, bring limitations to the instrument standardization? Future researches are suggested, with diagnosed persons as sample, in order to corroborate or compare the results discussed here.

References

- Amarani, R. (2009). Two Theories, One Theta: A Gentle Introduction to Item Response Theory as an Alternative to Classical Test Theory. *The International Journal of Educational and Psychological Assessment*, 3, 104-109.
- American Education Research Association (AERA), American Psychology Association [APA] & National Council on Measurement in Education (NCME). (1999). *Standards for Psychology and Educational Testing*. Washington, DC: American Psychology Association.
- American Psychiatric Association (APA). (2002). *DSM-IV-TR - Manual diagnóstico e estatístico de transtornos mentais*. (4ª ed). Porto Alegre: Artmed.
- Anastasi, A., & Urbina, S. (2000). *Testagem Psicológica*. Veronese, M.A.V. (trad.). Porto Alegre: Artmed Editora.
- Baptista, M. N. (2012). *Manual Técnico da Escala Baptista de Depressão - Versão Adulto (EBADEP-A)*. São Paulo: Vetor.
- Baptista, M. N., Cardoso, H. F., & Gomes, J. O. (2012). Escala Baptista de Depressão (Versão Adulto) - EBADEP-A: validade convergente e estabilidade temporal. *Psico-USF*, 17(3), 407-416.
- 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.
- Batistoni, S. S. T., Néri, A. N., & Cupertino, A. P. (2010). Validade e confiabilidade da versão Brasileira da Center for Epidemiological Scale - Depression (CES-D) em idosos Brasileiros. *Psico-USF*, 15(1), 13-22.
- Batistoni, S. S. T., Néri, A. N., & Cupertino, A. P. F. B. (2007). Validade da escala de depressão do Center for Epidemiological Studies entre idosos Brasileiros. *Revista de Saúde Pública*, 41(4), 598-605.
- Bauer, D. J., & Hussong, A. M. (2009). Psychometric Approaches for Developing Commensurate Measures Across Independent Studies: Traditional and New Models. *Psychological Methods*, 14(2), 101-125.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1997). *Terapia da Depressão*. Rio de Janeiro: Zahar.
- Calil, H. M., & Pires, M. L. N. (1998). Aspectos gerais das escalas de avaliação de depressão. *Revista de Psiquiatria Clínica*, 25(5), 240-244.
- Castro, S. M. J., Trentini, C., & Riboldi, J. (2010). Item response theory applied to the Beck Depression Inventory. *Revista Brasileira de Epidemiologia*, 13(3), 1-13.
- Cole, D. A., Martin, N. C., Youngstrom, E. A., Curry, J. F., Essex, M. J., Goodyer, I. et al. (2011). Structure and Measurement of Depression in Youths: Applying Item Response Theory to Clinical Data. *Psychological Assessment*, 23(4), 819-833.
- Cole, J. C., Smith, T. L., Rabin, A. S., & Kaufman, A. S. (2004). Development and Validation of a Rasch-Derived CES-D Short Form. *Psychological Assessment*, 16(4), 360-372.
- Conselho Federal de Psicologia (CFP) (2003). Resolução nº 002/2003 [On-line] Define e regulamenta o uso, a elaboração e a comercialização de testes psicológicos e revoga a Resolução CFP nº 025/2001. Retrieved from http://www.pol.org.br/legislacao/pdf/resolucao2003_2.pdf
- Covic, T., Pallant, J. F., Conaghan, P. G., & Tennant, A. (2007). A longitudinal evaluation of the Center for Epidemiologic Studies-Depression scale (CES-D) in a Rheumatoid Arthritis Population using Rasch Analysis. *Health and Quality of Life Outcomes*, 6, 6-41.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum.

- Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-381.
- Ferster, C. B., Culbertson, S., & Boren, C. P. (1977). *Princípios do comportamento*. (trad: Maria Ignez Rocha e Silva, Maria Alice de Campos Rodrigues e Maria Benedita Lima Pardo). São Paulo: Hucitec.
- Forkmann, T., Boecker, M., Wirtz, M., Eberle, N., Westhofen, M., Schauerte, P., et al. (2009). Development and validation of the Rasch-based Depression Screening (DESC) using Rasch analysis and structural equation modelling. *Journal of Behavior Therapy and Experimental Psychiatry*, 40(3), 468-478.
- Hambleton, R. K., & Jones, R. W. (1993). An NCME Instructional Module on Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hauck Filho, N., & Teixeira, M. A. P. (2011). A estrutura fatorial da escala CES-D em estudantes universitários brasileiros. *Avaliação Psicológica*, 10(1), 91-97.
- International Test Commission (ITC) (2000). Guidelines on Adapting Test. *International Test Commission*. Retrieved from <http://www.intestcom.org/upload/sitefiles/40.pdf>
- Jones, R. N., & Fonda, S. J. (2004). Use of an IRT-based latent variable model to link different forms of the CES-D from the Health and Retirement Study. *Social Psychiatry and Psychiatric Epidemiology*, 39, 828-835.
- Linacre, J.M. (2010). *Winsteps® (Version 3.70.0.2) [Computer Software]*. Beaverton, Oregon: Winsteps.com.
- Noronha, A. P. P. (2009). Testes Psicológicos: conceito, uso e formação do psicólogo. Em Claudio S. Hutz (Org.). *Avanços e polêmicas em avaliação psicológica*. São Paulo: Casa do Psicólogo.
- Noronha, A. P. P., & Alchieri, J. C. (2004). Conhecimento em Avaliação Psicológica. *Estudos de Psicologia*, 21(1), 43-52.
- Noronha, A. P. P., & Vendramini, C. M. M. (2003). Parâmetros psicométricos: estudo comparativo entre testes de inteligência e de personalidade. *Psicologia: Reflexão e Crítica*, 16(1), 177-182.
- Nunes, C. H. S. S., & Primi, R. (2009). Teoria de resposta ao item: conceitos e aplicações na psicologia e na educação. In Hutz, C S (org). *Avanços e polêmicas em avaliação psicológica*. São Paulo: Casa do Psicólogo.
- Organização Mundial da Saúde (OMS) (1993). *Classificação dos Transtornos Mentais e do comportamento - CID-10: descrições e diretrizes diagnósticas*. Trad. Dorgival Caetano. (3º Volume, 10ª Ed.). Porto Alegre: Artes Médicas.
- Pasquali, L. (1998). *Psicometria: Teoria e aplicações*. Brasília: Editora UnB.
- Pasquali, L. (1999). *Instrumentos psicológicos: manual prático de elaboração*. Brasília: LabPAM & IBAPP.
- Pasquali, L., & Primi, R. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica*, 2(2), 99-110.
- Pickard, A. S., Dalal, M. R., & Bushnell, D. M. (2006). A Comparison of Depressive Symptoms in Stroke and Primary Care: Applying Rasch Models to Evaluate the Center for Epidemiologic Studies Depression Scale. *Value in Health*, 1(9), 59-64.
- Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3), 385-401.
- Reise, S. P., & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, 5, 27-48.
- Rueda, F. J. M. (2007). O funcionamento diferencial do item no teste pictórico de memória. *Revista Avaliação Psicológica*, 6(2), 229-237.
- Santor, D. A., Gregus, M., & Welch, A. (2006). Eight Decades of Measurement in Depression. *Measurement*, 4(3), 135-155.
- Sauer, S., Ziegler, M., & Schmitt, M. (2012). Rasch analysis of a simplified Beck Depression Inventory *Personality and Individual Differences*, (in press). doi: 10.1016/j.paid.2012.10.025
- Silveira, D. X., & Jorge, M. R. (1998). Propriedades psicométricas da escala de rastreamento populacional para depressão CES-D em populações clínica e não-clínica de adolescentes e adultos jovens. *Revista de Psiquiatria Clínica*, 25(5), 251-61.
- Smith, A. B., Wright, E. P., Rush, R., Stark, D. P., Velikova, G., & Selby, P. J. (2006). Rasch analysis of the

- dimensional structure of the Hospital Anxiety and Depression Scale. *Psycho-Oncology*, 15, 817-827.
- Thomas, M. L. (2011). The Value of Item Response Theory in Clinical Assessment: A Review. *Assessment*, 18(3), 291-307.
- Urbina, S. (2004). *Essentials of Psychological Testing*. New Jersey: John Wiley & Sons, Inc.
- Valentini, F., & Laros, J. A. (2011). Teoria de Resposta ao Item na Avaliação Psicologia. In R. A. M. Ambiel, I. S. R., S. V. Pacanaro, G. A. S. Alves, I. F. A. S. Leme (Orgs). *Avaliação Psicológica: guia de consulta para estudantes e profissionais de psicologia*. São Paulo: Casa do Psicólogo.
- Wiberg, M. (2004). Classical Tests Theory vs. Item Response Theory: an evaluation of the theory test in the Swedish driving-license test. *Working Paper EM*, 50, 1-30.
- Wyse, A. E., & Reckase, M. D. (2011). A Graphical Approach to Evaluating Equating Using Test Characteristic Curves. *Applied Psychological Measurement*, 35(3), 217-234.

Received: September 5, 2013
Accepted: March 17, 2014