Validity Study of the Beck Anxiety Inventory (Portuguese version) by the Rasch Rating Scale Model

Estudo de Validade da Versão Portuguesa do Beck Anxiety Inventory Mediante o Modelo Rasch Rating Scale

Sónia Quintão[*, a], Ana R. Delgado[b] & Gerardo Prieto[b]

[a]Cedoc, Departamento de Saúde Mental, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal & [b]Universidade de Salamanca, Salamanca, Espanha

Abstract

Our objective was to conduct a validation study of the Portuguese version of the Beck Anxiety Inventory (BAI) by means of the Rasch Rating Scale Model, and then compare it with the most used scales of anxiety in Portugal. The sample consisted of 1,160 adults (427 men and 733 women), aged 18-82 years old (M=33.39; SD=11.85). Instruments were Beck Anxiety Inventory, State-Trait Anxiety Inventory and Zung Self-Rating Anxiety Scale. It was found that Beck Anxiety Inventory's system of four categories, the data-model fit, and people reliability were adequate. The measure can be considered as unidimensional. Gender and age-related differences were not a threat to the validity. BAI correlated significantly with other anxiety measures. In conclusion, BAI shows good psychometric quality.

Keywords: Anxiety, assessment, Beck Anxiety Inventory, Psychometrics, Rasch Rating Scale Model.

Resumo

Foi objetivo a validação da versão portuguesa do *Beck Anxiety Inventory* (BAI) mediante o modelo *Rasch Rating Scale* e a sua comparação com as escalas mais usadas de ansiedade em Portugal. A amostra consistia de 1160 adultos (427 homens e 733 mulheres) com idades entre 18-82 anos (M = 33,39; DP = 11,85). Os instrumentos utilizados foram: BAI, *State-Trait Anxiety Inventory* e *Zung Anxiety Scale*. Verificou-se que o sistema de quatro categorias, o ajuste dos dados ao modelo e a fidelidade das pessoas eram adequados. A medida é unidimensional. O género e as diferenças relacionadas com a idade não se mostraram ameaças à validade. O BAI correlaciona-se significativamente com as restantes medidas de ansiedade. Conclui-se que o instrumento apresenta boa qualidade psicométrica.

Palavras-chave: Ansiedade, avaliação, *Beck Anxiety Inventory*, Psicometria, *Rasch Rating Scale Model*.

Anxiety is a prevalent emotional disorder that interferes with psychosocial functioning (Balestrieri, Isola, Quartaroli, Roncolato, & Bellantuono, 2010). Thus, it is not surprising that most anxiety assessment tools have been developed in clinical settings.

Anxiety measuring instruments can be classified into those that assess only the neurovegetative components of the anxious response and the ones combining the evaluation of physiological components with the cognitive and behavioral components. The Beck anxiety inventory (BAI; Beck, Epstein, Brown, & Steer, 1988) is one of the most used clinical rating scales. In previous studies, BAI scores have shown high internal consistency, with Cronbach α of .92 and moderate test-retest reliability for one week with r = .75. BAI discriminated groups diagnosed as anxious (panic disorders, generalized anxiety, etc.) from groups

diagnosed as not anxious (major depression, atypical depression, etc.).

In the study of the Brazilian BAI version the scale had adequate reliability, with a Cronbach α of .91 for psychiatric samples, .86 for clinical samples, and .86 for non-clinical samples. The correlation between test and retest with a week of difference ranged from .53 for a sample of 115 students and .99 for a sample of 65 subjects of the general population (Cunha, 2001). Another study (Sanz & Navarro, 2003) examined the psychometric properties of a Spanish BAI version in a sample of 590 Spanish university students. BAI showed a high level of internal consistency, with a Cronbach α of .88 and factor analyses revealed a dimension formed by two very interrelated factors, corresponding to somatic and afective-cognitive symptoms. Taking the DSM-IV as the standard, the validity of BAI content was appropriate because their items covered 45% of symptomatic criteria specific of anxiety disorders and 78% of the symptoms of panic attacks.

Factor analyses have been conducted with results ranging from two to four factors (Beck et al., 1988; Beck

* Endereço para correspondência: Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Campo Mártires da Pátria, 130, Lisboa, Portugal 1169-056. E-mail: sonia.quintao@gmail.com

& Steer, 1990, 1991; Cox, Cohen, Direnfeld, & Swinson, 1996; Cunha, 2001; Hewitt & Norton, 1993; Steer, Ranieri, Beck, & Clark, 1993). However, BAI is usually treated as unidimensional whenever a total sum score is calculated.

For Leyfer, Ruberg and Woodruff-Borden (2006) BAI is not a diagnostic tool, but its brevity and simplicity make it an ideal instrument for use as a pretest for presence of anxiety disorder. The State-Trait anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970) is one of the self-assessment instruments most used internationally (Andrade & Gorenstein, 1998). In previous studies, Cronbach alpha have been found to range from .86 to .95 for the subscale STAI-State, and from .89 to .91 for the STAI-trait (Spielberger et al., 1970), whose scores have excellent test-retest reliability in multiple time intervals (Barnes, Harp, & Jung, 2002). Scores from the Zung Anxiety Scale (Zung, 1971) have also shown adequate internal consistency. Zung and BAI measure similar constructs, with emphasis on the somatic aspects of anxiety.

The objective of this study was to validate the BAI in Portugal with a modern psychometric model and then run a comparison of BAI, STAI trait, STAI State and Zung, the most used scales of anxiety in Portugal. The limitations of classical test theory, the usual model for construction and analysis of tests, has led to the emergence of alternative models, among which one of the most parsimonious is the Rasch model, which allows the conjoint measurement of persons and items (Bond & Fox, 2001; Rasch, 1960). A well-known extension of this model for polytomous data is the Rating Scale Model (Andrich, 1978; Prieto, Delgado, Perea, & Ladera, 2010; Stone, 2003). In order to fulfill our objective, we had to analyze the response categories, estimate the model parameters, their precision and degree of fit, test the scale dimensionality and the differential item functioning, and correlate the scores from BAI, trait STAI, State STAI and Zung.

Method

Participants

The sample was composed by 1160 adults from the Portuguese general population, 427 men and 733 women, with mean age of 33.39 (*SD*=11.85). Most subjects lived in urban areas (84.31%), were Caucasian (85.86%), college educated (68.19%), Catholic (65.26%), and employed (60.86%).

Instruments

We used a demographic questionnaire designed for this research, which asked about gender, age, residence, ethnicity, education level, religion and status, and the following anxiety instruments:

Beck Anxiety Inventory (BAI; Beck et al., 1988). It consists of 21 items, which are statements descriptive of anxiety symptoms that participants have to evaluate with reference to themselves, in a Likert scale of 4 points. The possible range of total scores goes from 0 to 63 (Beck et al., 1988; Cunha, 2001).

State-Trait Anxiety Inventory (STAI trait, STAI state; Spielberger et al., 1970). This questionnaire is composed of two blocks of 20 statements, evaluated in a four-point Likert scale. Form 1, STAI State, evaluates transient or temporary anxiety and form 2, STAI Trait, dispositional or general anxiety.

Zung Anxiety Scale (Zung, 1971). It was designed to assess situational anxiety. The scale consists of 20 statements evaluated in a four-point Likert scale. Scores range from a minimum of 20 to a maximum of 80. The 20 items are distributed in four anxiety subscales: Cognitive, Motor, Vegetative and Central nervous system, but only the total score was used in this study.

Procedure

Test Application Followed Ethical Standards. The implementation was carried out in various universities, companies and public facilities. Participants who did not comply with at least one item in BAI were removed from the database. Missing values were replaced by item averages. Reversed items were recoded. Data was analyzed with the program Winsteps, version 3.68 (Linacre, 2009).

Data Analysis

The model proposed by Rasch (1960) is based on two major assumptions: the attribute can be represented on a single dimension where people and items are conjointly located; and person level and item location are the only (probabilistic) predictors of a correct answer. The formula to model this relationship is:

$$\ln (P_{is} / 1 - P_{is}) = (B_s - D_i)$$

where B_s is the person parameter and D_i the item location.

With polytomous data, the formula for the Rating scale model is (Andrich, 1978):

$$\ln (P_{nik} / P_{ni(k-1)}) = B_n - D_i - F_k$$

where, P_{nik} is the probability that person n answer is category k;

$P_{ni(k-1)}$ is the probability that the observation or response is k-1;

B_n is the skill, attitude, trait… of person n;

D_i is the location of item i;

F_k is the transition point (step) between k and k-1.

This model is widely used in the analysis of scales with Likert format, in which all items are answered with the same set of ordered categories. The analysis of the functionality of the categories of response followed criteria proposed by Linacre (2002): (a) sufficient frequency and regular distribution of the chosen categories; (b) the average measures according to category should monotonically go up in the rating scale; (c) no category should show misfit, and (d) the transition points (steps) must increase monotonically.

Model fit (with pearsonian residual-based statistics) and score unidimensionality were then evaluated. Although strict unidimensionality is never achieved in practice (Zickar & Broadfoot, 2009), a principal component analysis of the residuals allows to assess whether the lack of unidimensionality is large enough to threaten score validity; the less stringent criterion is Reckase's (1979, cited in Zickar & Broadfoot, 2009), according to whom the percent of variance explained should be over 20% and there should not be a second dominant factor.

Differential Item Functioning (DIF) indicates lack of validity because the likelihood of an answer is determined by factors other than the construct measured. Currently, DIF analysis is an obligatory step in the validation of a test. Accordingly, we carried out DIF analyses with respect to gender and age (30 or less and more than 30). The procedure implemented in Winsteps estimates, for each item, the difference between item difficulty in each group (focal and reference). The contrast is carried out with the formula proposed by Wright and Panchapakesan (1969):

$$t = B_f - B_r / (SE^2_f + SE^2_r)^{1/2}$$

Where $B_f - B_r$ are item locations for the target and reference groups, and SE^2_f and SE^2_r are the squares of their typical errors. According to Wright and Douglas (1975), the DIF values that degrade the measures correspond to differences ($B_f - B_r$) over .5 logits. However, the Bonferroni correction is currently recommended to calculate a posteriori significant differences (Linacre, 2010).

Finally, factorial ANOVAs were carried out to test differences (impact) by sex and age in the Rasch-model scores. Previously we corroborated that assumptions for the use of parametric tests, i.e. normal distribution (Kolmogorov-Smirnov test) and homogeneity of variances (Levene test), were fullfiled.

Results

Every category system met Linacre's (2002) criteria as can be seen in Table 1. Once checked the adequacy of the categories, unidimensionality was put to the test. The BAI Rasch dimension, analogous to a first factor in a common factor analysis, explained 41.2 % of the variance: not optimal according to Linacre (2010), but still acceptable following Reckase (1979, cited in Zickar & Broadfoot, 2009). STAI-state, STAI-trait and Zung results were similar to BAI's, with 47.6%, 46.2%, and 38.9% variance explained, respectively. Thus, scores are essentially unidimensional.

Table 1
BAI, STAI-State, STAI-Trait and Zung Category Statistics

Category	Chosen category *f* and %	B	Infit	Outfit	Step
BAI					
1. Not at all	15498 (68%)	-2.88	1.03	1.02	--
2. Mildly	4721 (21%)	-1.61	.94	.79	-1.08
3. Moderately	2121 (9%)	-.70	1.04	1.08	-.37
4. Severely	403 (2%)	-.03	1.24	1.37	1.45
STAI-state					
1. Not at all	8151 (36%)	-2.43	.96	.97	--
2. Somewhat	9127 (40%)	-.92	.90	.82	-1.75
3. Moderately so	4090 (18%)	-.01	1.01	1.08	.31
4. Very much so	1530 (7%)	.79	1.20	1.38	1.44
STAI-trait					
1. Almost never	7601 (33%)	-2.51	-.99	.99	--
2. Sometimes	9903 (43%)	-1.12	.89	.86	-2.03
3. Often	4378 (19%)	-.05	.82	.83	.24
4. Almost always	1133 (5%)	.94	1.42	1.85	1.79
Zung					
1. A little of the time	11954 (52%)	-1.78	1.07	1.12	--
2. Some of the time	6637 (29%)	-.86	1.02	.82	-.75
3. Good part of the time	2765 (12%)	-.29	.96	.86	.25
4. Most of the time	1784 (8%)	.26	1.13	1.11	.50

As to model fit, no items were found exceeding 1.5 *Infit* and/or *Outfit*, excepting BAI item 16 (Fear of dying), STAI-state item 4 (Filling tired) and item 7 (Currently, I am concerned about possible woes), and Zung item 19 (I can only get a good rest during the night). Severe misfit was only found for STAI-trait item 24 (I wish I could be so happy as others seem to be) and Zung item 13 (I can inspire and expire with ease). The remaining items had values around unity (Linacre, 2009).

For the BAI, 9.31% of the participants show moderate misfit and 5.60% high misfit, in STAI-state 8.19% and 9.83%, in STAI-trait 8.19% and 8.71% and in Zung 9.91% and 11.21% respectively.

Item reliability was very high for every scale, close to 1.00. As to person reliability, BAI (.79) is reasonably good, STAI-state and STAI-trait are very good (.91 both) and Zung (.71) is moderate. These values have some similarity with the Cronbach's α of classical theory. Table 2 shows the summary of BAI results.

Table 2
Summary of the BAI Results

Statistics	Max.	Min.	*M*	*SD*	Value
Item *infit*	1.52	.71	1.07	.16	---
Item *outfit*	1.37	.66	.96	.20	---
% Items with moderate misfit (1)	---	---	---	---	4.76%(1)
% Items with severe misfit (2)	---	---	---	---	0%
Person *infit*	4.16	.26	.97	.42	---
Person *outfit*	5.40	.22	.96	.58	---
% Persons with moderate misfit (1)	---	---	---	---	9.31% (108)
% Persons with severe misfit (2)	---	---	---	---	5.60% (65)
Item difficulty	1.09	-1.69	.00	.79	---
Item reliability	---	---	---	---	.99
Person parameter	1.73	-4.46	-2.36	1.09	---
Person reliability	---	---	---	---	.79

Note. (1) *Infit* and/or *outfit* >1.5 and < 2; (2) *Infit* and/or *outfit* > 2; (3) *p*<.05.

Table 3 shows the BAI person-item conjoint representation. It can be seen that the person mean is much lower than the item mean, showing the low anxiety level of the sample.

No item showed DIF related to gender, and only two showed age-related DIF: STAI-trait item 32 and STAI-state item 18 (-.54 logits and -.65). These items did not work equally for participants below and over 30 even if they had the same level of anxiety. They should be excluded from the test if results are replicated in subsequent studies.

Factorial ANOVAs showed main effects of gender for all scales, with values between *F* (1; 1156) = 11.728; *p* < .001 (Zung) and *F* (1; 1156) = 21.466; *p* < .001 (STAI--trait), but only main effects of age in the BAI and the STAI-trait, with values of *F* (1; 1156) = 14.511; *p*< .001 and *F* (1; 1156) = 13.862; *p*< .001, respectively. No interaction effects were found. Male participants showed less anxiety in all scales, and participants older than 30 were less anxious.

Finally, correlations between BAI scores and the remaining anxiety measures were large and significant: *r* = .42, *p* <.001 (Zung), *r* = .55, *p* <.001 (STAI-trait), and *r* = .59, *p* <.001 (STAI state).

Discussion

Our main goal was to carry out an initial validation of the BAI for Portuguese population and to compare it with some other usually applied anxiety measures (STAI-state, STAI-trait and Zung). A psychometric model with optimal properties, the Rasch rating scale model, was used to test the functionality of the response category systems. This is seldom taken into account by the classical test theory in which determination of the categories is usually a priori. All evaluated scales showed good category functioning following Linacre's criteria (2002).

The Beck Anxiety Inventory is a scale with good psychometric characteristics, and in some contexts, such as the clinical one, in which the physiological symptoms are important, more appropriate than other scales used in Portugal.

Table 3

BAI Joint Representation of Persons (# is 8; '.' is 1 to 7) and Items

```
Scale      Persons        Items
                 . |
                   |T
                   |
                   |
                   | Item 19
 1             .  +  Item 9
                   |S Item 11; item 13
                 . |
                 . | Item 15; item 16; item 20; item 8
                 . | Item 12; item 3
                 . |
                 . | Item 21
 0             .  +M Item 1; item 17
             .# T|  Item 14; item 6
                 . |
               ## |
              .## | Item 18
               .# |
             .### |S Item 2
-1           .###  +  Item 7
              .## | Item 5
           .##### S|  Item 4
              .## |
          .####### |T
            .#### |  Item 10
            .#### |
-2          .#### +
           .##### |
        .######### |
       .######### M|
        .######### |
                   |
        .######### |
-3  .###########  +
                   |
       .########### |
                 S|
                   |
      .############ |
                   |
-4                 +
                   |
                   |
          .####### |
                 T|
                   |
                   |
-5      .######### +
```

BAI presented person reliability (similar to Cronbach α) reasonably good, but poor than the internal consistency presented in the original version (Beck et al., 1988) and in some countries like Brazil (Cunha, 2001) and Spain (Sanz & Navarro, 2003).

Although several studies point to the existence of more than one factor in the BAI (Beck & Steer, 1990, 1991; Cox et al., 1996; Steer et al., 1993), previously studied samples come from diverse populations, so that generalization is risky. From a practical point of view, a unidimensional measure makes sense when one of the factors is clearly dominant. Our analyses show that BAI, STAI-state, STAI--trait and Zung can be treated as unidimensional.

With just some exceptions, item-model fit was good enough. In BAI and STAI-state, no items with severe misfit were found. As regards severe person-model misfit, it was never over ten percent. Likewise, reliability estimates were high enough for every scale. It is worth noting that, although the BAI measures do not show higher reliability (Person Separation Reliability) than the other anxiety measures, this instrument presents the lowest total percentage of misfit item and the lowest percentage of severe misfit persons.

No item showed gender-related DIF and only two items from the STAI-trait and the STAI-state showed age-related DIF. As to impact, women had on average higher anxiety values, which is consistent with the scientific literature (Grillon, 2008). In relation to age, BAI, STAI-trait and Zung showed that the younger subsample had higher values of anxiety, results that are also consistent with past research (Spence, Rapee, McDonald, & Inaram, 2001).

Given that the instruments were originally designed to measure intensity of the anxiety symptoms, especially physiological symptoms (Beck et al., 1988; Leyfer et al., 2006; Spielberger et al., 1970; Zung, 1971), it is not surprising that most of the participants were below the mean range of the variable. It can be seen that person-item conjoint representation is a useful way of comparing anxiety levels and communicating results.

The Beck Anxiety Inventory is a measure widely used in international research, but is not used in Portugal for lack of evaluation of psychometric characteristics. In this study, the BAI showed a good evidence of validity and reliability.

The largest contribution of this research was to allow future research in Portugal to use the BAI as a tool for the evaluation of anxiety, as construct in general. This is of great importance, once that anxiety has been associated with an increased risk for other diseases, and plays an important role in the quality of life in general, as well as in relation to the capacity to drive in normal daily life. In addition, anxiety disorders involve high individual and social costs tend to be chronic and can be as disabling as somatic disorders (Lepine, 2002).

A limitation of this study was the fact that it wasn't used a clinical sample, being suggested for future studies the use of clinical samples, with medical or psychiatric disorders.

References

Andrade, L. H. S. G., & Gorenstein, C. (1998). Aspectos gerais das escalas de avaliação de ansiedade. *Revista de Psiquiatria Clínica, 67,* 285-290.

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Balestrieri, M., Isola, M., Quartaroli, M., Roncolato, M., & Bellantuono, C. (2010). Assessing mixed anxiety-depressive disorder. A national primary care survey. *Psychiatry Research, 176,* 197-201.

Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger state–trait anxiety inventory. *Educational and Psychological Measurement, 62,* 603-618.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety. *Journal of Consulting and Clinical Psychology, 56,* 893-897.

Beck, A. T., & Steer, R. A. (1990). *Manual for the Beck anxiety inventory*. San Antonio, TX: Psychological Corporation.

Beck, A. T., & Steer, R. A. (1991). Relationship between the Beck anxiety inventory and the Hamilton anxiety rating scale with anxious outpatients. *Journal of Anxiety Disorders, 5,* 213-223.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model*. Mahwah, NJ: LEA.

Cox, B. J., Cohen, E., Direnfeld, D. M., & Swinson, R. P. (1996). Does the Beck anxiety inventory measure anything beyond panic attack symptoms? *Behavior Research and Therapy, 34,* 949-961.

Cunha, J. A. (2001). *Manual da versão em português das Escalas Beck.* São Paulo, SP: Casa do Psicólogo.

Grillon, C. (2008). Greater sustained anxiety but not phasic fear in women compared to men. *Emotion, 3,* 410-413.

Hewitt, P. L., & Norton, G. R. (1993). The Beck anxiety inventory: A psychometric analysis. *Psychological Assessment, 5,* 408-412.

Lepine, J. P. (2002). The epidemiology of anxiety disorders: Prevalence and societal costs. *Journal of Clinical Psychiatry, 14*, 4-8.

Leyfer, O. T., Ruberg, J. L., & Woodruff-Borden, J. (2006). Examination of the validity of the Beck anxiety inventory and its factors as a screener for anxiety disorders. *Journal of Anxiety Disorders, 20,* 444-458.

Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement, 3,* 85-106.

Linacre, J. M. (2009). Winsteps (version 3.68) [Computer software]. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2010). *A user's guide to winsteps-ministep. Rasch-model computer programs*. Chicago, IL: Winsteps.com.

Prieto, G., Delgado, A. R., Perea, M. V., & Ladera, V. (2010). Scoring neuropsychological tests using the Rasch Model: An illustrative example with the Rey-Osterrieth Complex Figure. *The Clinical Neuropsychologist, 24,* 45-56.

Rasch, G. (1960). *Probabilistic Models for some intelligence and attainment tests.* Copenhagen, Denmark: Institute for Educational Research.

Sanz, J., & Navarro, M. E. (2003). Propriedades psicométricas de una versión española del inventario de ansiedad de Beck (BAI) en estudiantes universitarios. *Ansiedad y Estrés, 1,* 59-84.

Spence, S. H., Rapee, R., McDonald, C., & Ingram, M. (2001). The structure of anxiety symptoms among preschoolers. *Behavior Research and Therapy, 39*, 1293-1316.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the Stait-trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Steer, R. A., Ranieri, W. F., Beck, A. T., & Clark, D. A. (1993). Further evidence for the validity of the Beck anxiety inventory with psychiatric outpatients. *Journal of Anxiety Disorders, 7,* 195-205.

Stone, M. H. (2003). Substantive scale construction. *Journal of Applied Measurement, 4,* 282-297.

Wright, B. D., & Douglas, G. A. (1975). *Better procedures for sample-free item analysis* (Research Memorandum No. 20). Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago.

Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23-48.

Zickar, M. J., & Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 37-61). New York: Routledge.

Zung, W. (1971). A rating instrument for anxiety disorders. *Psychosomatics, 12,* 371-379.