

Psychometric Properties of the Inventário Dimensional Clínico da Personalidade (IDCP) using the Rating Scale Model

Propiedades psicométricas del Inventario Dimensional Clínico de Personalidad (IDCP) utilizando el Modelo Rating Scale
Propriedades psicométricas do Inventário Dimensional Clínico da Personalidade (IDCP) utilizando o Modelo Rating Scale

Lucas de Francisco Carvalho, Ricardo Primi*

Universidade São Francisco (USF)

Gregory E. Stone**

University of Toledo

Doi: [dx.doi.org/10.12804/apl32.03.2014.09](https://doi.org/10.12804/apl32.03.2014.09)

Abstract

The aim of this study was to evaluate the performance of the Dimensional Clinical Personality Inventory (DC-PI) using Rasch-based person and item analysis. 1281 participants were recruited, between 18 and 90 years of age ($M=26.64$; $SD=8.94$), 431 men (33.6%). Of the total sample, 127 (9.9%) were patients diagnosed with axis I disorders and/or axis II according to DSM-IV-TR. Results indicated the IDCP scales performed reasonably well, and the usefulness of the analyses presented, demonstrates the Rasch model's applicability for clinical applications. Among the important tools offered by the Rasch model, we explore the use of the person-item map, which visually presents the intuitively understandable psychological construct along the dimensional scale of the instrument.

Keywords: Item response theory; psychometric properties; personality disorders.

Resumen

El objetivo de este estudio fue evaluar el desempeño del Inventario Dimensional Clínico de la Personalidad (IDCP) utilizando el modelo Rating Scale para análisis de ítems y personas. Participaron 1281 sujetos entre 18 y 90 años de edad ($M = 26.64$, $DT = 8.94$), siendo 431 hombres (33.6%). De la muestra total, 127 (9.9%) eran pacientes diagnosticados con trastornos del Eje I y/o del Eje II según el DSM-IV-TR. Los resultados indicaron que las escalas del IDCP funcionan razonablemente bien, y la utilidad de los análisis presentados demuestra la aplicabilidad del modelo de Rasch para utilización clínica. Entre las herramientas más importantes que ofrece el modelo de Rasch, se explora el uso del person-item map, que presenta visual e intuitivamente la construcción psicológica comprensible a lo largo de la escala dimensional del instrumento.

* Lucas de Francisco Carvalho, Programa de pós-graduação stricto sensu em Psicologia, Universidade São Francisco (USF); Ricardo Primi, Programa de pós-graduação stricto sensu em Psicologia, Universidade São Francisco (USF).

** Gregory E. Stone, Professor do Educational Foundations and Leadership, University of Toledo.

Correspondence concerning this article should be addressed to Lucas de Francisco Carvalho, Universidade São Francisco, Rua Alexandre Rodrigues Barbosa, 45; CEP 13251-900, Itatiba; São Paulo, Brasil. Telephone number: 11 4534 8000. E-mail: lucas@labape.com.br

To cite this article: Carvalho, L., Primi, R., & Stone, G. E. (2014). Psychometric Properties of the Inventário Dimensional Clínico da Personalidade (IDCP) using the Rating Scale Model. *Avances en Psicología Latinoamericana*, 32(3), 433-446. doi: [dx.doi.org/10.12804/apl32.03.2014.09](https://doi.org/10.12804/apl32.03.2014.09)

Palabras clave: Teoría de respuesta al ítem; propiedades psicométricas; trastornos de la personalidad.

Resumo

O objetivo deste estudo foi avaliar o desempenho do Inventário Dimensional Clínico da Personalidade (IDCP) utilizando o modelo *Rating Scale* para a análise de itens e pessoas. Participaram 1.281 sujeitos entre 18 e 90 anos de idade ($M = 26,64$, $DT = 8,94$), sendo 431 homens (33,6%). Da amostra total, 127 (9,9%) eram pacientes diagnosticados com transtornos do Eixo I e/ou do Eixo II segundo o DSM-IV-TR. Os resultados indicaram que as escalas do IDCP funcionam razoavelmente bem, e a utilidade das análises apresentadas demonstram a aplicabilidade do modelo Rasch a utilização clínica. Entre as ferramentas mais importantes que oferece o modelo de Rasch, explora-se o uso do person-itemmap, que apresenta visual e intuitivamente a construção psicológica compreensível ao longo da escala dimensionada do instrumento.

Palavras chave: Teoria de resposta ao item; propriedades psicométricas; transtornos da personalidade.

According to Millon, Grossman, and Tringone (2010), personality disorders are different styles or patterns, i.e., characteristics sets that last over time and situations of pathological personality functioning. Grounded on the clinical background and diagnosis criteria of axis II Diagnostic and Statistical Manual of Mental Disorders IV-TR (DSM-IV-TR, APA, 2003), Millon (Davis, 1999; Grossman & Ramanath, 2004; Millon & Davis, 1996; Millon, Millon, Meagher, Millon, & Grossman, 2007a, 2007b; Strack & Millon, 2007) developed a integrative-evolutionary personality theory.

Based on the pathological characteristics of Millon's theory and axis II of DSM-IV-TR (DSM-IV-TR, APA, 2003), Carvalho (2011) developed the Inventário Dimensional Clínico da Personalidade (IDCP). Further empirical support for construction of the dimensional perspective was garnered from Schroder, Wormworth, & Livesley (1992). The IDCP is a self-report inventory, consisting of 163 items, distributed across 12 dimensions (for further

explanation about the dimensions, see Carvalho, 2011).

The 163 items of the IDCP were derived from an item bank consisting of over 500 items. 215 items were selected based on the theoretical conceptual point of view and were applied to over 1000 subjects, among non-patients and psychiatric patients. From this, it was set up a database and the data were subjected to various statistical analyzes, seeking to verify the internal structure of the instrument (composed of 12 distinct dimensions), validity evidences based on external criterion and reliability indexes (Carvalho, 2011).

Using IDCP, personality disorders may be evaluated in the 12 dimensions, aligned with pathological characteristics proposed by Millon (Millon & Grossman, 2007a, 2007b) and according to DSM-IV-TR (APA, 2003). In addition, the instrument is in line with the current trend for the future edition of DSM, the DSM 5, which is based on a dimensional diagnostics, considering that people should be assessed in all dimensions of personality, as a personality profile.

The IDCP was originally developed using Classical Test Theory (CTT). The assumptions of CTT create problems known in the social sciences as arbitrary metrics (Embretson, 2006). For example, CTT treats qualitative responses, ordinal in nature, as if they were immediately quantitative in the assignment of numeric representation. In doing so it fails to adequately capture respondent communication and provides inaccurate interpretations. Further, the lack of use of interval measures limits the generalizability of any inferences that may be made from information gleaned from the instrument. As the purpose of development is generally to create generalizable instruments useful outside the sample, such a limitation is severely confining.

Typically, psychological tests are interpreted with reference standards, which give meaning to test scores by comparing them to normative groups. Although the importance of such information is recognized, normative referencing neither establishes nor addresses the meaning of what is being measured *per se*, and therefore cannot reasonably explain changes in measures across the scale. In attempt to

address this issue, recent investigations have successfully made use of Item Response Theory (IRT) for developing and testing psychometric properties of tests for personality assessment, personality disorders and related constructs (Balsis, Gleason, Woods, & Oltmanns, 2007; Cooke & Michie, 1997; Feske, Kirisci, Tarter, & Pilkonis, 2007; Olatunji et al., 2009; Samuel Simms, Clark, Livesley, & Widiger, 2010; Stelmack et al., 2004; Walton, Roberts, Krueger, Blonigen, & Hicks, 2008). IRT models reflect a latent trait class approach that, unlike CTT, does not assume items are identical in scaling difficulty, and further defines item difficulty and person ability as functions of the probability of persons and items. IRT models fall into two categories: the Rasch Model and the 1-, 2-, or 3-parameter model. While the 1-parameter IRT model is mathematically equivalent to the Rasch model, Rasch considers person and item parameters to be both the only necessary but also fully sufficient statistics involved in the probabilistic function, and thus specifically excludes rather than holds as equal the discrimination parameter. The 2- and 3-parameter models add pseudo-guessing and discrimination parameters to the function, which better define the particulars of the sample but also add sample dependence and limit generalizability.

IRT offers direct and expeditious ways to establish diagnostic, clinically relevant standards. For example, the Item Reference Standard Setting Model (Embretson, 2000) used in this paper, allows for the assignment of meaning to scores on a test vis-à-vis standards (expected responses), allowing a more qualitative attribution of meaning on the numerical scale used (Carvalho & Primi, 2009; Carvalho & Primi, 2010; Linacre, 2009; Primi, 2004). As in the development of Item Reference Standard Setting use of IRT models offer key advantages to developers of instruments to measure psychological constructs.

Furthermore, the use of IRT models permits (a) an investigation of the structure and function of the categories used as a test responses (especially for Likert and/or rating scales), (b) a comparison of the intensity level of the construct represented in the items of a test with the intensity level of the

construct in persons (theta), (c) an investigation of the hierarchical organization of items according to the intensity represented by each of them, and (d) verification of the reliability of a test at the different levels at which the construct is measured. While there are certainly other advantages and application possibilities of IRT, an extensive survey is beyond the scope of this work.

There are several models based on IRT. One of the most frequently used is the Rasch model (Embretson, 2000). In the Rasch model, items are characterized only by the parameter b , called the level of difficulty, therefore this model has also been called the one-parameter IRT model. For the treatment of rating scales under conditions of the Rasch model (Wright, 1982), the Rating Scale Model was used. The rating scale model is an extended expression of the standard Rasch model. The standard Rasch model expresses the probability of a correct response as

$$\Pr\{X_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

where β_n is the ability of person n and δ_i is the difficulty of item i . The standard Rasch model is a dichotomy (0,1). In the Rasch rating scale model adjustments are made for the probability of respondents selecting one successive rating over the previous rating (i.e., thresholds). It is defined as follows:

$$\Pr\{X_{ni} = 1\} = \frac{\exp \sum_{k=0}^x (\beta_n - (\delta_i - \tau_k))}{\sum_{k=0}^m \exp \sum_{k=0}^x (\beta_n - (\delta_i - \tau_k))}$$

where δ_i is the difficulty of item i and τ_k is the k th threshold of the rating scale in common.

Considering the possibility of using IRT in the field of assessment of personality disorders, the aim of this study was to verify the parameters of the items and person for the IDCP obtained by the Rating Scale Model.

Method

Participants

Participants in this study included 1281 people, between 18 and 90 years of age ($M=26.64$; $SD=8.94$), and 61.8% ($N = 792$) were female. Of the 1281 participants, 1154 were undergraduate students of a town in the São Paulo state. The other 127 participants were patients of psychiatric clinics and the public hospital of São Paulo.

Instrument

In accordance with the objectives of this study the Inventário Dimensional Clínico da Personalidade (IDCP) was administered to all study participants. The IDCP is an instrument for assessing personality disorders based on Millon's theory and axis II of DSM-IV-TR (APA, 2003). It includes 162 items (15 items appear in more than one scale) divided into 12 distinct scales: Dependence (20 items related to the inability to trust yourself to make decisions depending on others for decision making.), Aggressiveness (27 items about reactions in which the individual does not consider the other to get what he desire, usually in a violent way), Humor Instability (27 items with respect to the tendency to sadness and irritable mood, but also to variations in mood, which often generate guilt), Eccentricity (20 items about the absence of pleasure in being with others, and beliefs that are different from other people, with manifested eccentric and idiosyncratic behaviors), Attention Seeking (16 items related to exaggerated need to get others attention, using mechanisms such as seduction, overreactions, and intensive search for friendships), Distrust (13 items respecting to persistent worry about being tricked, beliefs that there is always "ulterior motives", and preference for what is known, been persecutory), Grandiosity (12 items reporting irritability due to lack of recognition from others, showing an exaggerated need for admiration with underlying beliefs of entitlement and superiority), Isolation (11 items reporting a preference for being alone, irrita-

tion with the need of take orders from others, and decrease in pleasure with relationships), Criticism Avoidance (7 items about widespread beliefs of disability and, therefore that others will humiliate and criticize him), Self-Sacrifice (7 items related to an exaggerated disregard of self with clear trends to help others.), Conscientiousness (11 items about the need to do things in a more organized and orderly way as possible, with a focus on responsibility and obligations demonstrating excessive worry, perfectionism, and rigid rules in relationships), Impulsivity (5 items respecting to reactions of impulsivity and recklessness, with a taste for activities involving violence). Each item is answered using a 4-point rating scale ranging from 1, "has little to do with me", to 4, "has a lot to do with me". The estimated time for completion is approximately 30 minutes. The identification of a person's profile on the IDCP dimensions may suggest pathological functioning of the personality, which may resemble the typical profiles of personality disorders.

Previous validity evidence for the IDCP internal structure was reported for the twelve dimensions of the instrument by means of exploratory factor analysis and confirmatory factor analysis (Carvalho, 2011). In addition, the IDCP has demonstrated adequate levels of reliability (Cronbach's alpha greater than .70) for eleven of the twelve dimensions (Conscientiousness demonstrated an alpha equal to .69). Moreover, the IDCP dimensions correlated well with the dimensions and facets of the Brazilian version of the NEO Personality Inventory Revised ([NEO-PI-R]; Costa Jr. & McCrae, 2009) and psychiatric diagnoses. As a result, the relationships expected between the dimensions of the IDCP and psychiatric diagnoses of axis II of DSM-IV-TR (APA, 2003) are expected to be equivalent to the dimensions and facets of the NEO-PI-R.

Procedures

Prior to initiation, the proposed study was submitted to the Ethics Committee and was approved (Protocol number CAAE: 0350.0.142.000-08).

The instruments and the Informed Consent Form were administered to all participants. Only after agreeing to sign the form were participants able to participate in the study.

Participants in the study may have completed all or part of the instrument (whole instrument = 561, first half = 316, second half = 358). We adopted this procedure to enable the data collection on people who showed less available time participate on the research. The instrument was administered in the classrooms at the universities of São Paulo (private), Paraná (public) and Santa Catarina (private), and in the waiting rooms of private clinics and public hospitals of the state of São Paulo.

Data Analysis

After data were collected, statistical analyzes were performed to address the primary questions posed in the study. The collected data were analyzed using the Rasch model, specifically the Rating Scale Model, using the statistical software Winsteps (Linacre, 2009) verifying the parameters of the items and respondents.

One of the basic postulates of modeling via IRT is unidimensionality, that is, the model assumes that items measure a primary dimension and secondary dimensions have a negligible influence (Swaminatham & Hambleton, 1985). Thus, the unidimensionality verification of the IDCP dimensions was a necessary first step in the analysis. The specification of unidimensionality was verified using the Rasch principal contrasts analysis implemented through Winsteps, and the 2.0 eigenvalue criteria (Linacre, 2009), i.e., contrasts with eigenvalue greater than 2 were considered as a second dimension. To this end, we considered each factor of the IDCP as an independent, though related, scale.

Winsteps was used to calibrate the parameters of the items, implementing a method of maximum likelihood estimation (Joint Maximum Likelihood Estimation). To analyze the model fit, we considered the model fit indexes, infit and outfit. These indexes consist of average values of the residues (observed score – modeled score) standardized and squared, i.e., chi-square divided by degrees of freedom. Infit is more sensitive to items that are tar-

geted to the persons, while outfit is more sensitive to items that are far from the persons. Because of the idiosyncratic nature of the item patterns associated with persons, problematic infit patterns are frequently harder to diagnose and treat. As a result, outfit patterns, which tend to focus more carefully on responses, mistakes, and guessing, are often more useful from a practical perspective. Using the recommendations of the literature, we considered values above 1.3 and item-total correlations close to zero as indicative of misfit to the model (Linacre & Wright, 1994; Smith 1996; Wright & Linacre, 1994). In addition, values below .6 were considered as overfitting and redundant. Mean square was selected for use over the standardized due to the relatively moderate sample size. We also considered Rasch reliability indexes (based on internal consistency) and local error, response categories of the scales, quantitative and qualitative analyses of the person-items map, and the item map. We opted more for the use of the calculation of the local error rather than by calculating the curve information, considering that both provide similar information. Given the restraints of this paper, the analyses concerning the local error, response categories and person-items map will be provided for only one of the IDCP scales, Self-Sacrifice. It is worth noting that for purposes of analysis, the average difficulty of items (b) was set at zero.

Results and Discussion

This work aimed to evaluate the performance of the IDCP using the Rasch Rating Scale Model. The specification of unidimensionality was first verified through a Rasch principal contrasts analysis implemented through Winsteps. Using the performance indicators associated with the item and person parameters it is possible to calculate an expected response for each subject for each item. The discrepancy between the modeled response (expected) and the observed is the residue.

The principal contrasts analysis is performed on this new residue data matrix, based on the portion of responses not predicted by the model. Thus, if a contrast composed by a set of items with a magnitude greater than 2 (according to guidelines Linacre,

2009) appears, it suggests a second dimension that may potentially affect the data in order to confound the meaning of the first dimension. This analysis seeks to determine values of components with eigenvalues greater than or equal to 2.0. However, in the present study, none of the contrasts reached eigenvalues of 2.0 or greater. Once assured of the

unidimensionality of the scales, the analysis could be continued.

Table 1 presents descriptive statistics summarizing the latent trait (theta) of the respondents, their fit indexes (infit and outfit) and the number of items answered in each of the IDCP scales. In addition, this table summarizes the descriptive

Table 1
Person and items summarized descriptive statistics

		Person			<i>b</i>	Items			Reliability
		Theta	Infit	Outfit		Infit	Outfit	<i>R</i>	
Dependence	\bar{X} (SD)	-1.0 (1.0)	1.0 (0.5)	1.0 (0.5)	0 (0.4)	1.0 (0.1)	1.0 (0.2)		0.77 (0.81)
	Max.	2.1	3.3	4.1	0.7	1.3	1.4	0.4-0.6	1.84(2.03)
	Min.	-3.9	0.1	0	-0.8	0.7	0.7		
Aggressiveness	\bar{X} (SD)	-1.3 (0.8)	1.0 (0.4)	0.9 (0.6)	0 (0.4)	1.0 (0.1)	0.9 (0.2)		0.70 (0.73)
	Max.	2.0	4.9	9.9	1.1	1.4	1.5	0.3-0.5	1.51 (1.65)
	Min.	-3.7	0.2	0.2	-0.8	0.8	0.6		
Humor Instability	\bar{X} (SD)	-0.5 (0.9)	1.0 (0.4)	1.0 (0.5)	0 (0.4)	1.0 (0.1)	1.0 (0.1)		0.85 (0.87)
	Max.	2.7	2.8	3.8	0.8	1.4	1.3	0.4-0.6	2.34 (2.58)
	Min.	-4.0	0.1	0.1	-0.7	0.7	0.7		
Eccentricity	\bar{X} (SD)	-1.2 (0.9)	1.0 (0.4)	0.9 (0.5)	0 (0.3)	1.0 (0.1)	1.0 (0.2)		0.70 (0.73)
	Max.	2.4	2.8	3.3	0.5	1.3	1.3	0.4-0.6	1.53 (1.66)
	Min.	-3.5	0	0	-0.8	0.7	0.6		
Attention Seeking	\bar{X} (SD)	-0.3 (0.9)	1.0 (0.6)	1.0 (0.7)	0 (0.6)	1.0 (0.2)	1.0 (0.2)		0.72 (0.78)
	Max.	3.1	4.6	8.0	1.1	1.4	1.6	0.4-0.6	1.62 (1.88)
	Min.	-3.3	0	0	-1.1	0.5	0.6		
Distrust	\bar{X} (SD)	-0.4 (1.0)	0.9 (0.6)	0.9 (0.6)	0 (0.4)	1.0 (0.1)	1.0 (0.2)		0.69 (0.74)
	Max.	2.9	3.3	4.0	0.6	1.2	1.2	0.4-0.7	1.48 (1.67)
	Min.	-3.7	0	0	-0.7	0.7	0.7		
Grandiosity	\bar{X} (SD)	-0.8 (1.0)	1.0 (0.6)	1.0 (0.6)	0 (0.5)	1.0 (0.1)	1.0 (0.1)		0.64 (0.70)
	Max.	2.7	4.2	5.0	0.7	1.2	1.2	0.5-0.6	1.33 (1.52)
	Min.	-3.5	0	0	-1.1	0.8	0.7		
Isolation	\bar{X} (SD)	-0.7 (0.9)	1.0 (0.6)	1.0 (0.6)	0 (0.3)	1.0 (0.1)	1.0 (0.1)		0.60 (0.66)
	Max.	2.5	3.3	5.0	0.6	1.2	1.2	0.5-0.6	1.23 (1.40)
	Min.	-3.2	0	0	-0.5	0.7	0.7		

Continúa

		Person				Items			Reliability
		Theta	Infit	Outfit	<i>b</i>	Infit	Outfit	<i>R</i>	
Criticism Avoidance	\bar{X} (SD)	-1.3 (1.1)	1.0 (0.6)	0.9 (0.6)	0 (0.4)	1.0 (0.2)	0.9 (0.2)		0.60 (0.66)
	Max.	2.8	3.4	3.8	0.6	1.3	1.4	0.6-0.7	1.23 (1.38)
	Min.	-2.9	0	0	-1.0	0.7	0.6		
Self-Sacrifice	\bar{X} (SD)	-0.4 (1.3)	0.9 (0.7)	0.9 (0.8)	0 (0.5)	1.0 (0.1)	0.9 (0.1)		0.63 (0.71)
	Max.	3.4	4.7	4.8	0.9	1.1	1.1	0.7-0.7	1.30 (1.55)
	Min.	-3.5	0	0	-0.6	0.9	0.9		
Conscientiousness	\bar{X} (SD)	0.1 (0.7)	0.9 (0.6)	1.0 (0.6)	0.1 (0.7)	1.0 (0.2)	1.0 (0.2)		0.52 (0.61)
	Max.	2.6	4.4	6.7	2.6	1.3	1.3	0.4-0.6	1.04 (1.24)
	Min.	-2.8	0	0	-2.8	0.7	0.7		
Impulsivity	\bar{X} (SD)	-0.8 (0.9)	0.9 (0.6)	0.9 (0.7)	0 (0.3)	1.0 (0)	0.9 (0)		0.29 (0.39)
	Max.	2.3	3.7	4.7	0.5	1.1	1.0	0.6-0.7	0.64 (0.79)
	Min.	-2.4	0	0	-0.6	0.9	0.9		

data for the items (i.e., the difficulty level, the fit indexes, the correlation item-theta, and reliability indices - real and modeled - and separation -real and modeled

In general, the average levels of the latent traits suggest that the items tended not be endorsed by the sample, except for the scale Conscientiousness, where theta showed a positive theta average. The scales with the lowest mean theta (-1.37 and -1.35) were the Aggressiveness and Criticism Avoidance, respectively, indicating that the items of these scales were the least endorsed by the participants. Although the average level in the latent trait of participants was low, the observed range of scores on all scales suggests that the sample is composed of people with both healthier, and more pathologic personality characteristics. The Rasch model allows this intuitive inference to be made that the scores of the subject, mild or more extreme, is indicative of the level of personality functioning. Altogether, 12 items have been found with some misfit in outfit or Infit statistics, ranging from 3 items in dimensions Aggressiveness and Humor Instability and no item in the dimensions Distrust, Grandiosity, Isolation,

Impulsivity and Self-Sacrifice. The low frequency of items with misfit is also suggestive of unidimensionality.

Also in relation to the participants, through the fit indexes, infit and outfit, there were detected discrepancies between the observed and expected values with respect to the estimation of thetas. These values tended to be acceptable (Linacre & Wright, 1994), because the mean value was below 1.3 for all scales. However, the fit indexes maximum values were higher than 1.3, suggesting discrepancies for some subjects according to what is expected by the model. The model explained more correctly more than 70% of the subjects for all scales. Moreover, the reliability index of theta estimates calculated by the Rasch model ranged between 0.29 and 0.85 (real) and 0.39 and 0.87 (modeled). These indices may be considered ranging from poor to satisfactory, particularly because some scales have a small number of items, and because the average level of difficulty of the items and the average level of subjects in the latent trait demonstrate wide ranges. Both characteristics can influence the calculation of reliability indices (Embretson, 2000).

With respect to the items descriptive data, the difficulty index varied between -2.86 and 2.62 on the Conscientiousness scale. The mean items fit indexes of all scales were adequate (less than 1.3), although some scales showed maximum scores that reached 1.3 or more. Also, the item-theta correlations indicated high positive correlations between the items and their dimensions, which also suggest cohesion between the components (items) for each dimension. Complementing the information about the reliability of dimensions, we also calculated the local error.

One of the advantages of using IRT is to understand the conditioned reliability to each scale (i.e., to know in which level of the scale the instrument has a higher reliability rate.) This is done by evaluating the local error curve that presents available information across the levels of theta. One way to express a standardized curve ranging from 0 to 1 is through the local error (Daniel, 1999).

This index allows for the assessment of which levels of theta (latent trait) of items (and IDCP scales) is more error-free (i.e., more reliable). For example, a scale with a moderate reliability may

be highly reliable in a certain range of latent trait, but less so at other levels. It should be noted that for calculating the local error, we considered only 477 subjects, specifically, those who responded to most of the items of each scale. The criterion for selection was the number of respondents in the Humor Instability scale, which had the lowest number of responding to all items. Figure 1 shows the reliability indices for the Self-Sacrifice scale in accordance with the level of the theta (local error).

In figure 1, the x-axis (horizontal) refers to the theta (ranging between -5 and +5) and the y-axis to the reliability indices. The horizontal line that cuts the graph is dividing the curve in reliability indices equal to or greater than 0.80, and indices below this cutoff. From there, one can check in which range of theta the Self-Sacrifice scale is more reliable. This range includes values of theta between -2.22 and 2.14, and the average reliability in this range is 0.88 (between 0.80 and 0.90). This finding contrasts with the “general” reliability of this dimension (0.71), since the weighting for different latent trait levels can increase or decrease.

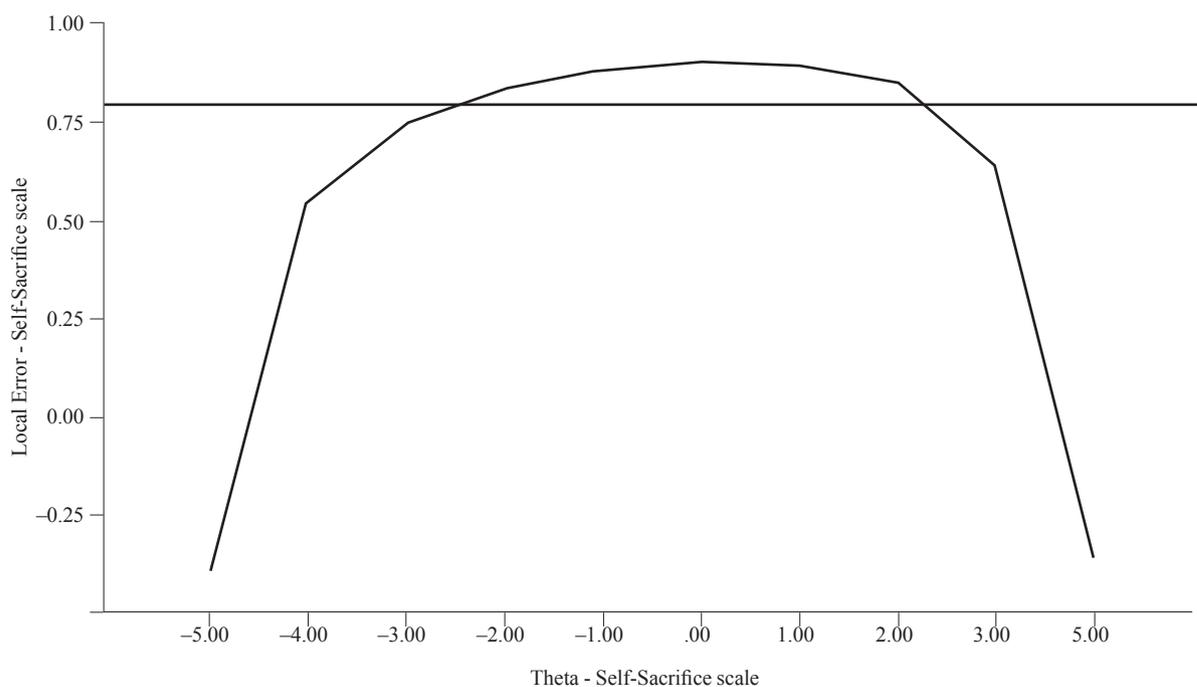


Figure 1. Local error – Self-Sacrifice scale

As expected, the reliability index of the Self-Sacrifice scale is higher for higher levels in the latent trait, since IDCP is focused on pathological personality functioning. There is no space for presentation of this information on all scales, but this same pattern of reliability increases within certain ranges of theta was observed for all scales of the instrument. These data suggest that the dimensions of the IDCP more appropriately evaluate pathological levels of personality functioning. Figure 2 provides illustrative data about the response categories of Self-Sacrifice scale.

Theta (x-axis) is paired with the response probability of participants at different levels of theta (y-axis) to describe each of the rating scale options. In the figure, the average b is centered on zero. Thus, it is possible to verify the likelihood of endorsement of the participants in each category of response and their distributions in different levels of theta for an item $b_i = 0$ (i.e., the average level of difficulty equal to zero). The four response categories ranged from (1) “has nothing to do with me”, (2) “has little to do with me”, (3), “has to do with me”, to (4) “has a lot to do with me”. The intersection between two categories can be interpreted as the threshold value of transition between these categories. The

threshold between the first and second categories is equal to -1.65, between 2 and 3 equal to 0.28, and between 3 and 4 equal to 1.37. A clear representation of all categories was observed (i.e., the curves do not overlap in at least one theta range.) Separation of the curves in different regions of the theta scale is a desirable metric feature because it indicates that respondent demonstrate clear differentiation between each rating scale category, and the present empirical data shows that the response to stimuli (items) has been quantitatively modeled by means of a increasing monotonic relationship between theta and categories. The response categories were appropriate according to the criteria presented earlier across all IDCP dimensions. The thresholds of the categories of response were also observed and in all cases were found that the theta increases monotonically to the ratings progress for all dimensions of the IDCP.

Figure 3 presents one of the most important applications of IRT to psychiatric disorders assessment, the person-item map, using the example of the Self-Sacrifice scale once again. As pointed out in this work, with IRT it is possible to employ item (criterion) referenced standard setting (Embretson, 2000), allowing one to assign meaning

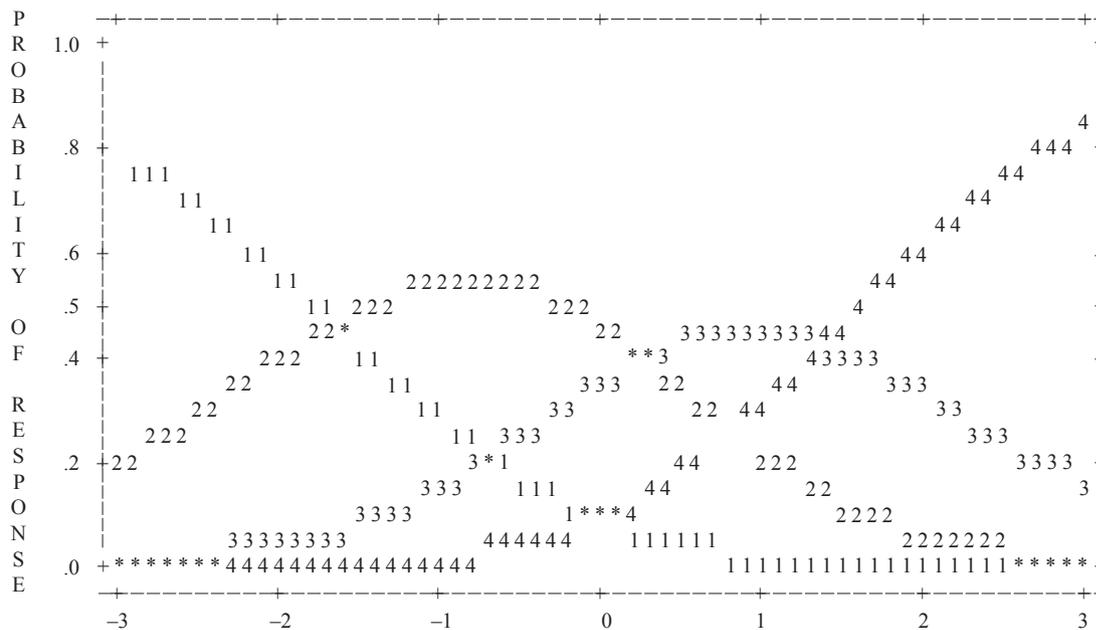


Figure 2. Self-Sacrifice scale response categories

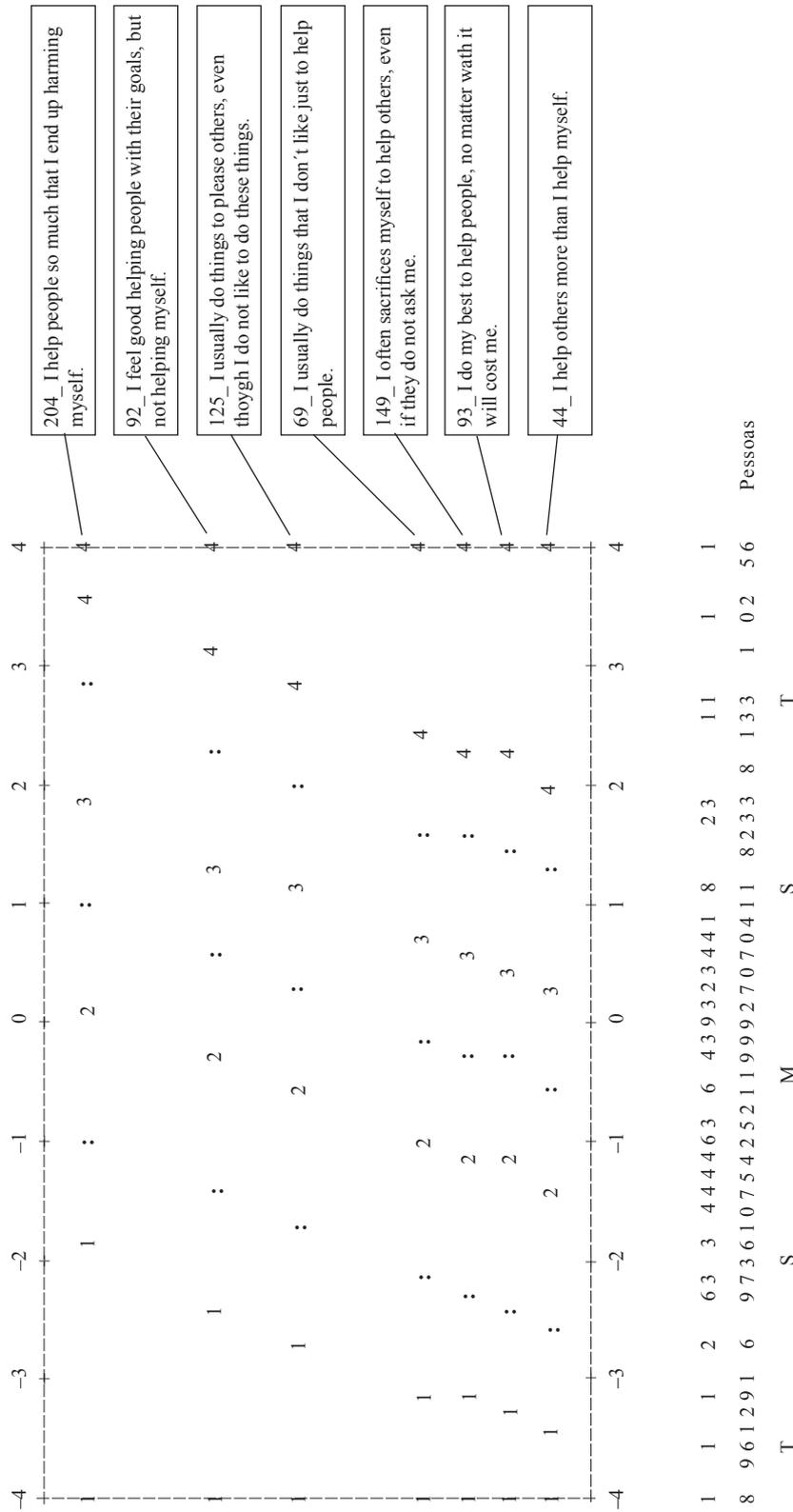


Figure 3. Self-Sacrifice scale Person-items map

to the scores of respondents at different levels of scale. The items are presented, from the bottom up, starting with the most endorsed to least endorsed ones. The number and content of each item can also be observed. The response categories (1-4) can be verified in the figure for each item of the dimension.

At the bottom of the figure is shown the distribution of respondents (number of responders in each theta level must be read vertically) and theta range (ranging from -4 to +4). Just below the distribution of the participants are letters T, S, M, which refer to, respectively, two standard deviations (T = above or below the average), one standard deviation (S = above or below the average), and mean (M). For this study, a qualitative analysis was used for the items of Self-Sacrifice scale considering the theoretical perspective underlying the construct in an attempt to bring clinical contributions of the items composing the scale.

A higher concentration of responders could be found between the theta range varying from -2.0 to 1.0, which was expected according to the average theta observed (see table 1). Moreover, there was a greater proportion of respondents in the lower theta categories of the sample, since most of the respondents had no psychiatric diagnosis. Overall, the content of the items concerned more or less directly with the exaggerated disregard of self and over consideration to others, as well as reactions of help and sacrifices for others with harm to the self, featuring central to the masochistic personality disorder (Millon & Grossman, 2007a).

The hierarchical arrangement of the items suggests that items 44, 93, 149 and 69, with content relating to the focus on helping others, but still mild, tended to be easier for the participants to endorse. The next item on the hierarchy, 125, namely, helping others even when one do not want to do that, is very specific. In sequence, item 92 seems more difficult than the earlier to endorse, probably in consideration of the fact that the person feels good when helping others, but in contrast, there is no good feeling when the person helps him or herself (i.e., displeasure in helping one's self). Lastly, item number 204, presents more intense content in a continuum of a health-pathologic perspective, in which the person claims to help others bringing

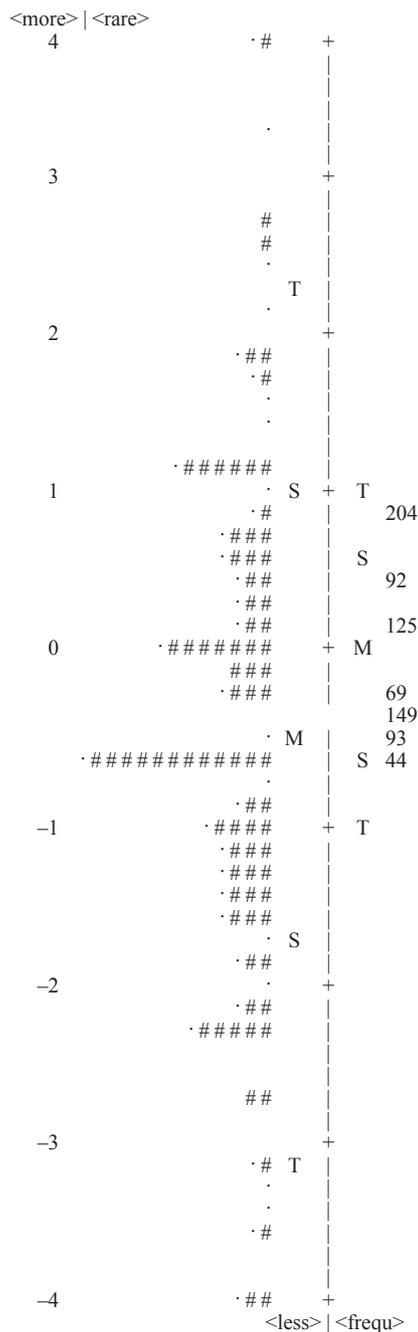
harm to herself. Therefore, it is possible to verify that the extent that decreases the endorsement of participants, i.e., the items becomes more difficult, the more the item content relates to personality pathological functioning (Millon & Grossman, 2007a).

It is interesting to note how the classical and item referenced standard setting procedures are complementary, allowing for a better understanding of the scale reference points. Note that the selection of categories 3 or 4 ("has to do with me" or "has a lot to do with me" respectively) on item 69, pathological elements are more evident, which corresponds to theta levels slightly above average.

The presented analysis demonstrates that persons with certain levels of the latent trait (in this case, characteristics related to masochist functioning) tend to agree with some of the statements, in a less likely progressive fashion. For example, people with theta equal to -0.5 tend to agree only with the first item (upwards), while people with theta equal to zero tend to agree with the first 4 items. This difference of 0.5 between the two levels of theta, used as examples, point to substantial changes in the personality functioning of these people. Thus, the standardized scalar index, theta, is not an arbitrary number on the scale. Instead it is possible to infer which features are present or not in a person with a certain level in the latent trait (Embretson, 2006).

Similar data are shown in figure 4, item map. To the left is shown the distribution of the sample and the right side the distribution of items. Whereas most of the sample has no known psychiatric diagnosis, it is expected that items tend to be less endorsed. Each "#" represents 13 people and each "." 1 to 12 people. Most of the people is located in a bottom range, with items showing a greater mean. The hierarchical items order is the same as the figure 3.

The item map also allows to verify the representativeness of the construct, by the items, in relation to the sample. For example, in Self-Sacrifice dimension, gaps in less severe levels and more severe levels are observed. Since the IDCP is to assess the pathological functioning, items assessing less severe levels are not required, but there is a need for the addition of more severe items for this dimension.



Similarly, the Distrust, Grandiosity, Isolation and Conscientiousness dimensions also showed a need to the insertion of more pathological items.

Conclusions

This study aimed to evaluate the item and person parameters and instrument functioning obtai-

ned by the Rasch model of the IDCP. Overall, the results suggest the adequacy of the psychometric properties of the scales of the instrument. Among the contributions of IRT to clinical instrument development, the person-item map should be emphasized, because it focuses clinical understanding of the scores obtained by individuals who respond to a particular group of items on a continuum of latent trait development. It is also worth noting the use of local error in addition to the reliability analyses conventionally used, offer the ability to check for different reliability indices that may vary across levels of the latent trait measured by the items.

Among the limitations of the study, two should be highlighted. First, the number of psychiatric cases used in the sample was relatively small (N=127), and IDCP is focused on pathological personality traits. Further, certain scales of the IDCP, such as Impulsivity, include few items, making dimensionality analyses difficult. In future studies, one should focus on psychiatric cases in the sample composition, and should also seek to develop more items for some scales in an attempt to assess more broadly the typical characteristics of different personality functioning. Thus, we hoped that this research contributes to the field of assessment of personality disorders, especially in light of modern psychometric procedures, which are already being used widely in other countries in the field of personality studies.

References

American Psychological Association. (2003). *Manual Diagnóstico e Estatístico de Transtornos Mentais DSM-IV-TR* (4ª ed.). Porto Alegre: Artmed.

Balsis, S., Gleason, M. E. J., Woods C. M., & Oltmanns T.F. (2007). Age group bias in DSM-IV personality disorder criteria: An item response theory analysis. *Psychological Aging*, 22, 171-85.

Carvalho, L. F. (2008). *Construção de um Instrumento para avaliação dos Transtornos da Personalidade*. (Dissertação de mestrado não publicada, Universidade São Francisco, Itatiba).

Carvalho, L. F. (2011). *Desenvolvimento e Verificação das Propriedades Psicométricas do Inventário Dimensional Clínico da Personalidade*. (Tese

- de doutorado não publicada, Universidade São Francisco, Itatiba).
- Carvalho, L. F., Bartholomeu, D., & Silva, M. C. R. (2010). Instrumentos para Avaliação dos Transtornos da Personalidade no Brasil. *Avaliação Psicológica*, 289-298.
- Carvalho, L. F., & Primi, R. (2010). *Development of a Brazilian Inventory for the Assessment of Personality Disorders Based on Millon's Model*. Painel apresentado na Society for Personality Assessment Annual Meeting, California, EUA.
- Carvalho L. F., & Primi R. (2009). *Personality Style Assessment in Patients with Chronic pain*. Painel apresentado na Society for Personality Assessment Annual Meeting, Chicago, EUA.
- Cooke, D. J., & Michie C. (1997). An item response theory analysis of the Hare Psychopathy Checklist--Revised. *Psychological Assessment*, 9(1), 3-14.
- Costa Jr., P. T., & McCrae, R. R. (2009). *NEO-PI-R - Inventário de Personalidade NEO Revisado - Manual*. São Paulo: Vetor.
- Craig R. J., & Bivens, A. (1998). Factor structure of the MCMI-III. *Journal of Personality Assessment*, 70, 190-96.
- Daniel, M. H. (1999). Behind the scenes: using new measurement methods on the DAS and KAIT. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum.
- Davis, R. D. (1999). Millon: Essentials of his science, theory, classification, assessment, and therapy. *Journal of Personality Assessment*, 72(3), 330-352.
- Dyce, J. A., O'Connor, B. P., Parkins, S., & Janzen, H. (1997). Correlational structure of the MCMI-III personality disorder scales and comparison with other data sets. *Journal of Personality Assessment*, 69(3), 568-82.
- Embretson S. E., & Reise S.P. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum.
- Embretson, S. E. (2006). The Continued Search for Nonarbitrary Metrics in Psychology. *American Psychologist*, 61(1), 50-55.
- Feske, U., Kirisci, L., Tarter, R.E., & Pilkonis, P.A. (2007). An application of item response theory to the DSM-III-R criteria for borderline personality disorder. *Journal of Personality Disorders*, 21, 418-33.
- Hambleton, H. K., & Swaminatham, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer.
- Handler L., & Meyer, G. J. (1997). The importance of teaching and learning personality assessment. Em Handler, L., & Hilsenroth, M. *Teaching and learning personality assessment*. New Jersey: Lawrence Erlbaum Associates.
- Linacre, J. M. (2009). *WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch analysis (Computer Software)*. Chicago, Illinois: MESA Press.
- Linacre, J. M., & Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(2), 370.
- Millon, T., & Davis, R. D. (1996). *Disorders of Personality DSM-IV and Beyond*. New Jersey: Wiley.
- Millon, T., & Grossman, S. (2007a). *Moderating severe personality disorders*. New Jersey: John Wiley & Sons Inc.
- Millon, T., & Grossman, S. (2007b). *Overcoming resistant personality disorders*. New Jersey: John Wiley & Sons Inc.
- Millon, T., Grossman, S., & Tringone, R. (2010). The Millon Personality Spectrometer: a tool for personality spectrum analyses, diagnoses, and treatments. Em: Millon, T., Krueger, R. F., & Simonsen. (Orgs.), *Contemporary directions in psychopathology: scientific foundations of the DSM-V and ICD-11*, (p. 610) New York: The Guilford Press.
- Millon T., Millon C. M., Davis R. D. (1994). *MCMI-III Manual*. Minneapolis: Dicandrien.
- Millon, T., Millon, C. M., Meagher, S. Grossman, S., & Ramanath, R. (2004). *Personality Disorders in Modern Life*. New Jersey: Wiley.
- Morana, H. C. P. (2003). *Identificação do ponto de corte para a escala PCL-R (Psychopathy Checklist Revised) em população forense brasileira: caracterização de dois subtipos da personalidade; transtorno global e parcial*. (Tese de doutorado não publicada, Universidade de São Paulo, São Paulo).
- Olatunji, B. O., Woods, C., Jong, P. J., Teachman, B., Sawchuk, C. N., & David, B. (2009). Development and initial validation of an abbreviated Spider

- Phobia Questionnaire using item response theory. *Behavior Therapy*, 40, 114-30.
- Pasquali, L., & Primi, R. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica*, 2, 99-110.
- Pasquali, L., & Primi, R. (2007). Fundamentos da Teoria de Resposta ao Item – TRI. Em Pasquali, L. *Teoria de Resposta ao Item: Teoria, Procedimentos e Aplicações*. Brasília: LabPAM/Unb.
- Primi, R. (2004). Avanços na Interpretação de Escalas com a Aplicação da Teoria de Resposta ao Item. *Avaliação Psicológica*, 3(1), 53-58.
- Rossi, G., Brande, I., Tobac, A., Sloore, H., & Hauben, C. (2003). Convergent validity of the MCMI-III personality disorder scales and the MMPI-2 scales. *Journal of Personality Disorders*, 17(4), 330-340.
- Rossi G., Van der Ark, L. A., & Sloore, H. (2007). Factor analysis of the Dutch-language version of the MCMI-III. *Journal of Personality Assessment*, 88, 144-57.
- Samuel, D., Simms, L.J., Clark, L.A., Livesley, J., & Widiger, T.A. (2010). An item response theory integration of normal and abnormal personality scales. *Personality Disorders: Theory, Research, and Treatment*, 1, 5-21.
- Schroeder, M. L., Wormworth, J. A., & Livesley, W. J. (1992). Dimensions of personality disorder and their relationships to the Big Five dimensions of personality, *Psychological Assessment*, 4(1), 47-53.
- Smith R. M. (1996) Polytomous Mean-Square Fit Statistics. *Rasch Measurement Transactions*, 10(3), 516-517.
- Stelmack, J., Szlyk, J. P., Stelmack, T., Babcock-Parziale, J., Demers-Turco, P., Williams, T. R., & Massof, R. W. (2004). Use of Rasch person-item map in exploratory data analysis: A clinical perspective. *Journal of Rehabilitation Research and Development*, 41(2), 233-241.
- Strack, S., & Millon, T. (2007). Contributions to the dimensional assessment of personality disorders using Millon's model and the Millon Clinical Multiaxial Inventory (MCMI9-III). *Journal of Personality Assessment*, 89(1), 56-69.
- Walton, K.E., Roberts, B.W., Krueger, R.F., Blonigen, D.M., & Hicks, B.M. (2008). Capturing abnormal personality with normal personality inventories: An item response theory approach. *Journal of Personality*, 76, 1623-1647.
- Widiger, T. A., & Trull, T. J. (2007). Place Tectonics in the Classification of Personality Disorder: shifting to a dimensional model. *American Psychologist*, 62(2), 71-83.
- Wright B. D., & Linacre J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA.

Received: September 26, 2013
Accepted: June 10, 2014