Psychological Thought
psyct.swu.bg | 2193-7281
South-West University "Neofit Rilski"

Research Article

Item Banking for an Adaptive Measurement of Neuroticism

Facundo Juan Pablo Abal*[a] , Gabriela Susana Lozzia[b], Sofía Esmeralda Auné[c], Horacio Félix Attorresi[d]

[a] University of Buenos Aires, National Scientific and Technical Research Council (CONICET), Argentina.

[b]University of Buenos Aires, Argentina.

[c] University of Buenos Aires, National Scientific and Technical Research Council (CONICET), Argentina.

[d]University of Buenos Aires, Argentina.

Abstract

The psychometric properties of a bank of 36 items are presented measuring Neuroticism based on the Five-Factor Model. These items pertain to the facets that were identified by the work of McCrae and Costa. The sample was comprised of 1133 adult subjects that reside in the Buenos Aires Metropolitan Area in Argentina. Women accounted for 52.1% of those subjects with an average age of 29.5 years (*SD* = 11.32). In order to get the items calibrated according to Item Response Theory (Graded Response Model), acquire the bank's information functions and assess the estimated associations with other instruments, 70% of the cases were randomly selected. An adaptive administration simulation was made with the remaining 30% so as to test two stopping rules: a) using 18 items and b) standard error of ≤ 0.25. Correlations greater than .95 were found between the estimated bank scores and the two adaptive versions. The advantages of using the adaptive Neuroticism measurement over other well-renowned instruments that use conventional large formats, as well as abbreviated ones, are discussed.

Keywords: neuroticism, Five Factor Model, item bank, computerized adaptive test, Item Response Theory.

Table of Contents

Psychological Thought, 2020, Vol. 13(2), 459-485, https://doi.org/10.37708/psyct.v13i2.503
Received: 2020-06-04. Accepted: 2020-08-16. Published (VoR): 2020-10-31.
Handling Editor: Natasha Angelova, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria.
*Corresponding author at: University of Buenos Aires. National Scientific and Technical Research Council (CONICET), Argentina. E-mail: afjp79@hotmail.com

Within the framework of the Five Factor Model (FFM), Neuroticism can be defined as a trait that describes the tendency to experience negative emotions such as fear, sadness, guilt and anger (Goldberg, 1993; McCrae & Costa, 2010). Individuals with high Neuroticism levels experience these emotions intensely and for unusually long periods, often leading to a state of prolonged ill-tempered states. They are prone to be highly dissatisfied with themselves and with the context, whereby they also tend to have difficult relationships (McCrae & Costa, 2010). Problems regulating these emotions negatively impact their ability to make decisions, think clearly and cope effectively with stress (Barlow et al., 2014; Widiger, 2009).

These features describing this trait do not necessarily imply the presence of pathology, since Neuroticism is a domain of normal personality. Nevertheless, the empirical evidence shows a profuse impact of Neuroticism on health conditions (Frølund-Pedersen et al., 2016; Jeronimus et al., 2016; Lahey, 2009; Sauer-Zavala et al., 2017; Vittengl, 2017). Subjects with high Neuroticism levels are likely to exaggerate the importance of physical symptoms, to use health services more frequently (Hajek et al., 2017; ten Have et al., 2005), and to use more medication with or without a prescription (Chapman & Goldberg, 2017).

Given that Neuroticism is a vulnerability factor for the development and maintenance of diverse physical illnesses and psychopathological disorders (Lahey, 2009), it is also recognized as an important variable for proposing intervention and prevention strategies from a transdiagnostic approach (Widiger & Oltmanss, 2017). Current studies state that Neuroticism is more malleable than it was previously assumed. It is for this reason that the efficacy of certain treatments have begun to be tested in recent years focusing on the objective to reduce Neuroticism (e.g. Drake et al., 2017; Sauer-Zavala, et al., 2017) and also, to recommend the detection of high levels of this trait in the general population during routine clinical care (Hengartner et al., 2016; Tackett & Lahey, 2017; Widiger, 2009).

The measurement of Neuroticism for screening purposes in the clinical context, as with other personality traits, requires efficient tests that allow a valid and reliable measurement in the shortest possible time. A similar demand arises in large-scale epidemiological assessments, wherein a reduction in items could be used to measure other variables of interest (Baldasaro et al., 2013). The most recognized inventories are usually extensive because they define a hierarchical model of Neuroticism composed of facets and use a considerable amount of elements to make an exhaustive evaluation of each of these sub-dimensions (Goldberg et al., 2006; McCrae & Costa, 2010). There are also shorter scales that perform a one-dimensional assessment of the domain (Goldberg, 1992; McCrae & Costa, 2007) or reduce the number of facets by either eliminating them (Soto & John, 2017b) or by subsuming them (DeYoung et al., 2007). Even short and extra short tests have been developed (Donnellan et al., 2006; Gosling et al., 2003; Soto & John, 2017a) in order to reduce measurement errors caused by fatigue or boredom.

However, the practical gain provided by these short or abbreviated forms is achieved at the expense of resigning psychometric quality (Credé et al., 2012). A brief scale that evaluates a broad construct such as Neuroticism, includes elements with moderate correlations that reflect the relative heterogeneity of the content and, consequently, decrease the internal consistency indices considered to examine reliability (Baldasaro et al., 2013; Sibley, 2012). The most frequent solution has been to select the items with greater discriminative capacity to elevate the internal consistency, neglecting the representativeness and exhaustiveness of the content (Milojev et al., 2013; Morizot, 2014; Ziegler et al., 2014).

The advance of modern psychometry has made it possible to apply the developments of Item Response Theory (IRT) to make the evaluation of personality traits more flexible and efficient by means of Item Banks and Computerized Adaptive Testing (CAT) (Attorresi et al., 2009; Reise & Revicki, 2015). In a CAT, an algorithm is used to progressively choose the items in a bank that provides more information which is based on the responses of the subject. The administration continues until either the standard error drops below a specified level, or the participant has answered the maximum number of questions. This adaptive procedure has an important practical advantage since it would allow the evaluation of Neuroticism to be shortened without compromising the reliability or measurement validity, as is the case with traditional forms of administration.

Item Banks and CAT have begun to have greater visibility in the context of instrumental studies in order to evaluate personological aspects (e.g. Abal et al., 2019; Nieto, et al., 2017; Rubio et al., 2007; Stark et al., 2012). Their construction is more expensive than a

conventional test, but they have demonstrated important practical advantages (Reise & Revicki, 2015). Indeed, Clinical and Health Psychology are the areas in which the development of CAT has been encouraged for the detection of pathological levels of Depression and Anxiety (e.g. Beiser et al., 2016; Devine et al., 2016; Gibbons et al., 2016).

The general objective of this study was to introduce the construction process of a bank of items for the measurement of the Neuroticism domain according to McCrae and Costa's Five Factor Model (2010). The aim was to contribute with an assessment tool appropriate to the characteristics of the local population, with discriminatory capacity based on individual differences, with solid evidence of validity, and with reliability studies of the generated measurements. In light of the instrumental demands pointed out to obtain an optimal measurement of Neuroticism, and considering the most recent psychometric advances, the following objectives were proposed for this study: a) to calibrate a set of Neuroticism items with IRT to build a bank, b) to obtain evidence of validity based on the correlation between Neuroticism and the variables of personality and psychological symptomatology, c) to examine if the adaptive administration of these items could show advantages and d) to analyze if adaptive administration affects the correlation of Neuroticism with external criteria.

Drawing from these objectives, the following hypotheses were formulated:

H1) Despite Neuroticism being defined as a personality domain made up of facets, the construct is expected to fit a unidimensional model to guarantee the bank modeling using IRT.

H2) The score estimated from the Neuroticism item bank will be significantly associated with its conceptually-related personality and psychopathological variables.

H3) Adaptive administration allows the instrument to be shortened without compromising its measuring quality.

Method

Participants

The sample consisted of 1133 general-population adults residing in the metropolitan area of Buenos Aires, Argentina who chose to collaborate. The subjects were selected from a non probability sampling method (convenience sampling). 52.1% of them considered themselves female. The mean age of all participants was 29.5 years (*SD* = 11.32; *Min* = 18, *Max* = 82). The majority of participants (57.2%) completed secondary studies while 15.5% had tertiary

studies and 22.2% had a university degree. Only 5.1% did not complete secondary education.

Instruments

Neuroticism Item Bank. A compilation of items was made from multiple instruments that evaluate both Neuroticism and its six facets and other related features. The contents collected from these empirical indicators were used as sources for the elaboration of new items that adjust to the composition of facets proposed by McCrae & Costa (2010). According to these authors, the definition of each facet is based on some kind of negative emotion or feeling that provides it with entity. The facets Anxiety and Hostility are built upon emotions of fear and anger respectively, while Depression and Self-consciousness are based on the feelings of sadness and shame. Impulsivity and Vulnerability, on the other hand, respond to a behavioral order. While the former is described as the impossibility of resisting temptations, the latter is characterized by the difficulty of implementing effective coping strategies in stressful situations.

A primary depuration was carried out based on the criticism of seven expert judges and a pilot study, which allowed the selection of the 36 items (see Appendix) that were being administered (six items for each of the facets). Of these, six items belonged to the Argentine adaptation of the IPIP-NEO Inventory (Cupani et al., 2014) that was included in the International Personality Item Pool (IPIP) by Goldberg et al. (2006). It was the experts' opinion that all the items were congruent with the conceptual definition of the facet that they operationalize (Aiken's $V \geq .85$). In order to avoid the violation of the IRT local dependency assumption, a qualitative analysis of the selected items' content was carried out, which allowed to verify that they were not mutually redundant (Abal et al., 2010; Reise & Rodriguez, 2016).

All items had a Likert response format of four options (*Disagree*, *Slightly Disagree*, *Slightly Agree* and *Agree*). This decision was based on recommendations derived from empirical and simulation studies (e.g. Abal et al., 2018; Lozano et al., 2008) wherein four categories were found to be an optimal amount to ensure a balance between the degree of the IRT model fit and the measurement reliability.

Eysenck Personality Questionary Revised short version, EPQ-R (Eysenck & Eysenck, 1994; adapted from Squillace et al., 2013). It comprised 42 items with a dichotomous response. At the local level, the adapters replicated the three-factor structure of the Eysenck model

(Psychoticism, Extraversion and Neuroticism) and the fourth Lie factor. The reliability studies of the four scales recorded adequate internal consistency indices (KR-20 between .66 and .84), which were slightly lower than those obtained with the sample of the present study (KR-20 between .69 and .86).

Symptoms Checklist – 90 Revised SCL-90-R (Derogatis, 1994). It consisted of 90 items that were grouped to enable the measurement of the intensity of the symptomatology perceived using a seven-day time reference in nine clinical dimensions (Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation and Psychoticism). It also permitted the obtention of three global indices: Global Severity Index, Positive Symptom Distress Index, Positive Symptom Total. The items had a five-option response format (from 0 – *not at all*, to 4 – *extremely*). The local adaptation showed validity evidence and reliability studies suitable for both non-clinical (Casullo, 2004) and clinical populations (Sánchez & Ledesma, 2009). The internal consistency of all items in the inventory showed a Cronbach's alpha of .96 (total scale, 90 ítems) in the sample of this study, while this coefficient ranged between .77 (Hostility scale, 6 items) and .86 (Depression scale, 13 items) for clinical dimensions.

Procedure

Individuals responded to the protocol individually, without any time limit and using a paper-and-pencil format. The administrations were carried out by psychologists, duly trained and supervised so that they carried out the applications in a correct evaluation environment according to what is commonly expected.

The examinees were informed about the purpose of this study. Before its administration it was explained to them that the task consisted in responding to a series of inventories that sought to evaluate personality features. It was emphasized that there were no correct or incorrect answers to the questions and that dedication and sincerity in answering was desired. They were informed about the voluntary nature of their participation and the possibility of abandoning the evaluation at any time during the activity. They were also notified that the anonymity and confidentiality of their responses were guaranteed. These considerations were detailed in writing and formed part of the consent that the subjects had to sign before responding.

Data analysis

Participants were randomly divided into two subsamples. Responses from 70% of the subjects *(n* = 793) were used to calibrate items with Samejima's Graded Response Model

(GRM). The rest of the individuals (n = 340) were considered exclusively for analyzing the efficiency of the CAT.

Item Calibration. A Confirmatory Factorial Analysis (CFA) was performed using the Mplus program (Muthén & Muthén, 2010) in order to verify the GRM assumption of unidimensionality. The parameters were estimated using the Weighted Least Squares Mean and Variance Adjusted (WLSMV) method on the basis of the polychoric correlation matrix. To confirm the degree of fit, the indicators and criteria recommended by Byrne (2012) were considered: Comparative-Fit-Index (CFI) and Tucker-Lewis-Index (TLI) greater than .90 and a Root Mean Square Error of Approximation (RMSEA) less than .08.

The GRM item calibration was performed using the MULTILOG program (Thissen, 2003). Marginal Maximum likelihood procedure was used to estimate item and response parameters. For each of the 36 items there was an estimation of a discrimination parameter (a) and three location parameters (b_1, b_2, b_3) that separate the adjacent categories of the Likert scale. In addition, a parameter θ was estimated to quantify their trait level per every subject. The GRM fit was studied by MODFIT (Stark, 2007). This program provided graphs that allowed the comparison of observed and expected probabilities for each item response category at 25 trait levels determined by default. Thus, the program offered information to define whether the model adequately predicts empirical curves. The fit was also assessed with the χ^2 index dividing MODFIT degrees of freedom (χ^2/df) from the comparison of pairs and triads of items. Following Drasgow et al. (1995), it was considered that ratio values of χ^2/df over 3 reflected problems regarding model-fit.

*Item Bank reliability and validity studies***.** Global reliability indicators were obtained: Cronbach's alpha, ordinal alpha and marginal reliability (Thissen, 2003). The Test Information Function (TIF) and the standard error of measurement that was found were plotted. Additionally, evidence of convergent and discriminant validity was obtained considering the correlations between the N estimates made with the complete bank and the EPQ-R and SCL-90-R scales.

Adaptive algorithm. The adaptive administration was studied with the Firestar software (Choi, 2009). A post hoc simulation was performed using the data matrix of the 340 separate sample cases that were intended for this purpose. This procedure consisted in the algorithm progressively choosing the items it would present in the case of an evaluatee responding to a CAT. Then, it retrieved the stored responses of the subject to choose the next item.

The trait mean was used as the initial estimate of the θ at the beginning of the administration. Successive provisional estimates from θ were made with the Bayesian method of Expected A Posteriori (EAP) measure using the normal standard as a priori distribution. The selection of items was made using the Maximum Fisher Information Selection Criterion, which allowed progressively selecting the most informative items for each provisional θ estimated, from the pool of items that have not yet been presented. In order to achieve a greater representativeness of the content in the sampling of the items, it was specified that the selection should be made at random among the three items with maximum information. Finally, two stopping criteria were tested: a) fixed length when administering 18 items (equivalent to 50% of the bank) and b) variable length when a target measurement precision has been attained (standard error of ≤ 0.25, equal to a classical reliability of .94).

The efficiency of both procedures was analyzed by correlating the θ estimated from the CAT (in its different stopping rules) with the θ estimated by responding to all items. The impact of adaptive measurement on the relationship of θ with the EPQ-R and SCL-90-R scales was also examined.

Results

Item Calibration

Unidimensionality. The results showed that the data properly fitted the undimensional model *CFI*= .91, *TLI*= .90, *RMSEA*=.055, 90 *CI* [.053 - .056]. The factor loading of items is shown in table 1. This evidence reasonably fulfilled the unidimensional assumption as required by the *GRM*.

GRM application. Fifty-eight iterations were required to reach the convergence criterion of the estimation parameters. Table 1 shows the item parameters and the standard error of estimation put in order according to the facet that operationalizes them: Anxiety (items 1-6), Hostility (items 7-12), Depression (items 13-18), Self-consciousness (items 19-23), Impulsivity (items 24-29) and Vulnerability (items 30-36). The location parameters b_1, b_2 and b_3 of the set of items were located along the different levels of the trait, mainly between -3 and 3. The appearance of a *b* parameter out of range was associated with items whose trait description were extreme. The *a* parameters showed, on average, moderate values with a mean of 1.18 (*SD* = 0.38, *Min* = 0.70, *Max* = 2). The comparison of the *a* parameters, in accordance with the content of the items, revealed variations associated with the facets, with Impulsivity and Hostility being the ones that take lower values.

Table 1.

Factorial loadings, estimated parameters according to IRT and percentage of use of adaptive version items.

Item	Factorial loading	Graded Response Model Parameters				% of item use in CAT	
		a (se)	*b_1 (se)*	*b_2 (se)*	*b_3 (se)*	Fixed length	Variable length
1	.44	0.80 (0.10)	-1.77 (0.24)	-0.62 (0.14)	1.09 (0.19)	13.3	9.5
2	.67	1.54 (0.12)	-0.73 (0.09)	-0.15 (0.08)	0.87 (0.08)	92.3	79.0
3	.59	1.28 (0.11)	-1.43 (0.14)	-0.36 (0.09)	0.84 (0.11)	76.9	44.1
4	.64	1.44 (0.12)	-1.47 (0.13)	-0.33 (0.08)	0.75 (0.08)	86.4	73.7
5	.72	1.82 (0.18)	-0.89 (0.08)	0.12 (0.07)	1.17 (0.10)	95.0	88.5
6	.75	2.00 (0.11)	-0.80 (0.07)	0.03 (0.06)	0.93 (0.08)	93.5	86.1
7	.43	0.78 (0.08)	-1.45 (0.22)	0.06 (0.13)	1.92 (0.26)	9.2	6.5
8	.39	0.70 (0.06)	-3.70 (0.56)	-2.00(0.33)	0.34 (0.17)	1.2	2.4
9	.39	0.71 (0.07)	-3.33 (0.50)	-1.21(0.22)	0.93 (0.21)	1.8	2.4
10	.61	1.28 (0.12)	0.00 (0.09)	1.13 (0.12)	2.43 (0.22)	60.9	26.0
11	.50	0.97 (0.11)	0.45 (0.11)	1.52 (0.19)	2.71 (0.32)	15.4	4.1
12	.40	0.72 (0.09)	0.93 (0.19)	2.31 (0.37)	3.81 (0.61)	2.1	0.9
13	.62	1.42 (0.09)	-1.14 (0.11)	-0.17(0.08)	1.04 (0.11)	90.8	70.4
14	.71	1.76 (0.11)	-0.88 (0.09)	0.05 (0.07)	1.11 (0.10)	92.6	84.3
15	.67	1.63 (0.21)	0.22 (0.07)	1.06 (0.10)	1.92 (0.16)	68.9	50.0
16	.60	1.33 (0.09)	-1.18 (0.12)	0.69 (0.10)	1.88 (0.17)	74.3	41.1
17	.68	1.61 (0.13)	-0.90 (0.09)	0.10 (0.07)	1.13 (0.10)	92.3	86.7
18	.64	1.46 (0.12)	-0.32 (0.08)	0.55 (0.08)	1.63 (0.14)	89.1	71.3
19	.48	0.85 (0.08)	-1.46 (0.21)	-0.28 (0.13)	1.23 (0.19)	15.7	12.4
20	.59	1.24 (0.11)	0.25 (0.09)	1.10 (0.12)	1.95 (0.19)	47.3	18.9
21	.54	1.08 (0.10)	-0.24 (0.11)	0.78 (0.12)	2.05 (0.23)	60.7	12.1
22	.43	0.80 (0.08)	-0.33 (0.14)	1.20 (0.19)	3.15 (0.41)	1.2	3.0
23	.57	1.15 (0.10)	-0.77 (0.12)	0.40 (0.10)	1.79 (0.18)	72.2	30.8
24	.64	1.34 (0.11)	-0.85 (0.11)	0.05 (0.08)	1.15 (0.12)	84.9	63.9
25	.39	0.71 (0.09)	-2.61 (0.18)	-0.47 (0.18)	1.60 (0.29)	0.3	2.4
26	.60	1.32 (0.12)	-0.45 (0.09)	0.50 (0.09)	1.65 (0.16)	82.2	57.1
27	.46	0.87 (0.09)	-0.14 (0.12)	0.80 (0.15)	2.24 (0.29)	2.1	4.4
28	.41	0.73 (0.08)	-0.62 (0.16)	1.24 (0.21)	3.06 (0.44)	0.6	4.1
29	.64	1.42 (0.12)	-0.26 (0.08)	0.60 (0.09)	1.60 (0.15)	81.1	62.4
30	.43	0.79 (0.08)	-0.06 (0.13)	1.41 (0.21)	3.00 (0.40)	2.1	1.8
31	.40	0.77 (0.07)	-0.94 (0.18)	0.82 (0.17)	2.33 (0.32)	7.1	4.4
32	.47	0.97 (0.11)	-0.78 (0.14)	0.62 (0.13)	2.45 (0.29)	34.6	16.6
33	.71	1.89 (0.11)	-0.07 (0.06)	0.92 (0.08)	1.91 (0.14)	83.1	69.5
34	.50	1.00 (0.08)	-0.65 (0.13)	0.69 (0.12)	2.13 (0.24)	46.4	16.9
35	.59	1.29 (0.12)	-0.17 (0.09)	1.12 (0.12)	2.36 (0.23)	65.7	27.8
36	.53	1.04 (0.09)	-0.99 (0.14)	-0.15 (0.10)	0.78 (0.13)	56.8	20.7

Note. se = Standard error.

The MODFIT fit graphs showed that all of the items' characteristic curves remained within the confidence interval related with the observed probability for each of the 25 contrasted trait levels. In the same line, the goodness-of-fit indices summarized in table 2 showed that all the χ^2/df ratios were less than 3 in the comparisons, when evaluating the item dyads and triad. In

conclusion, both the graphical methods and the fit indices manifested that the GRM was adequate so as to model Neuroticism items. The set of 36 calibrated items formed the Item Bank that was used in the next phase of the CAT study.

Table 2.

Frequencies and descriptive statistics of χ^2/df ratio to evaluate the fit to the Graded Response Model.

	Frequency Distribution of Adjusted χ^2 to *df* Ratio			*M*	*SD*
	<1	1<2	2<3		
Singles	36	0	0	0.05	0.03
Doubles	15	17	4	1.31	0.58
Triplets	3	9	0	1.31	0.31

Bank reliability and validity studies. Every global reliability coefficient showed highly satisfactory values (Cronbach alpha = .92, ordinal alpha = .95 and marginal reliability = .93). The FIT was relatively symmetrical with respect to θ = 0 demonstrating that the bank was reliable to measure the Neuroticism level where the largest number of individuals were located (Figure 1).

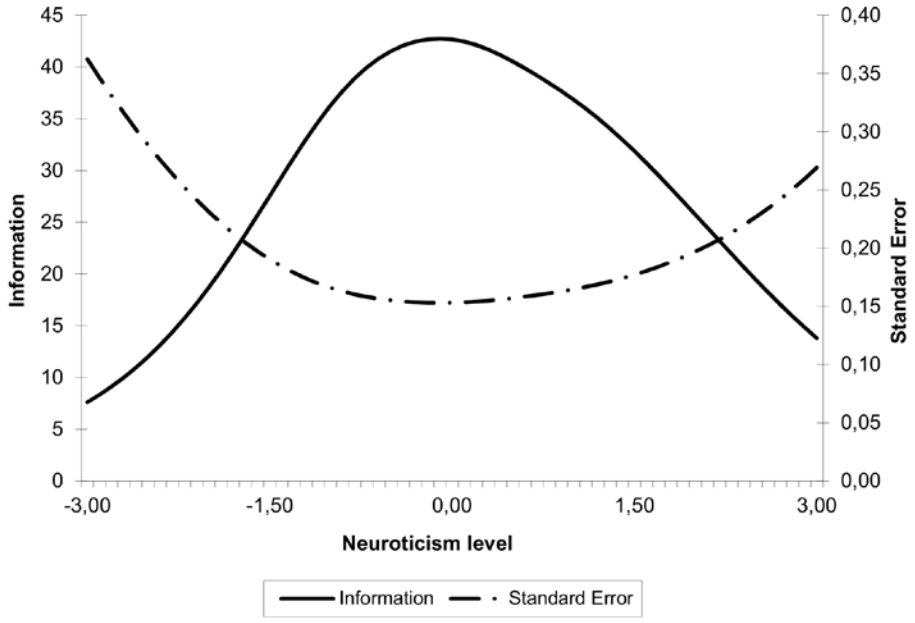

Figure 1. Test Information and Standard Error Functions

The associations between the θ estimates with the complete bank and the EPQ-R and SCL-90-R scales showed results according to what was expected from a theoretical perspective (Table 3). The θ correlated moderate-high with the Neuroticism scale of the EPQ-R and low with the rest of the variables measured by this questionnaire. Likewise, all correlations with

SCL-90-R were positive and moderate. The intensities of these associations varied according to the conceptual closeness of Neuroticism to the different symptom patterns evaluated by SCL-90-R.

Table 3.

Correlations of complete bank estimates and CAT with other constructs.

Instruments	Scale	Complete Bank (*df* = 791)	CAT Fixed length (*df* = 338)	CAT Variable length (*df* = 338)
Complete Bank		1**	.971**	.951**
Classical Score		.99**	.944**	.921**
EPQ-R	Neuroticism	.776**	.761**	.744**
	Extraversion	.204**	.189**	.182**
	Psychoticism	.038	.039	.021
	Lie	.162**	.144**	.158**
SCL90-R	Somatization	.431**	398**	450**
	Obsessive -Compuls	.526**	516**	523**
	Interpersonal Sensitivity	.609**	.575**	.568**
	Depression	.600**	.592**	.628**
	Anxiety	.678**	.658**	.660**
	Hostility	.489**	.443**	.505**
	Phobic Anxiety	.357**	.339**	.378**
	Paranoid Ideation	.491**	.454**	.499**
	Psychoticism	.533**	.492**	.517**
	Global Severity Index	.700**	.696**	.703**
	Positive Symptom Total	.555**	.533**	.522**
	Positive Symptom Distress Index	.399**	.354**	.426**

Note. $^{**}p < .001$

CAT simulation

A high and positive correlation was found between the θ estimates with the bank and both its fixed $r(338) = .98$, $p < .001$ and variable $r(338) = .95$, $p < .001$ length adaptive version. The intensity of these correlations decreased when considering the association of the CAT with the total score calculated with classical theory (Table 3).

The standard error of estimation at θ obtained when responding to the fixed-length CAT varied between 0.18 and 0.31 with an average of 0.22 (*SD* = 0.023). This implied that an optimum level of precision was reached, even when the number of items that were administered were reduced by half. Under the conditions established by the variable length version, it was required to administer an average of 12.6 items (*SD* = 4.41) per subject. After presenting 12 items, 59.4% of participants reached an error ≤ 0.25 and 91.7% required 18 items or less. Only two people (0.59%) did not reach the pre-established error and their

evaluation was interrupted due to reaching the 36-item limit. Both evaluatees adopted scores θ located over 1.5 standard deviations below the trait mean and the standard error did not exceed .27.

Table 1 shows the percentage of cases in which each item was administered in adaptive versions. In relation to the starting rule method (trait mean) and item selection method (Maximum Fisher Information), the most used bank items were those whose *b* parameters were located close to θ = 0 and which had high and moderate *a* parameters. Those contents linked to Hostility were unlikely to be chosen (items 7 to 12) because the low values of their *a* parameters reduced the chances of them being administered. This problem was worsened in the variable length CAT since Hostility items presented lower use percentages than in the fixed length CAT.

Finally, in Table 3, the correlations between the estimated θ with the CAT and the EPQ-R and SCL-90-R variables can be seen. The correlation indices found showed that, in general, the Neuroticism relations with other external variables were not altered. This suggested that the tested version of the adaptive administration of items did not substantially impact the evidence of convergent and discriminant validity.

Discussion

The clinical relevance of Neuroticism that has been demonstrated in recent years makes it crucial to think of evaluation instruments that adjust to the demands of application contexts in which efficient measurement is prioritized. The theoretical consolidation of the FFM has provided enough motivation to create short instruments for Neuroticism and the other personality domains that have been found to be useful for evaluating large samples (e.g. Cupani & Lorenzo-Seva, 2016; Donellan, et al., 2006; Gosling et al., 2003; Natividade & Hutz, 2015; Soto & John, 2017a, 2017b). But the strategies used by the Classical Test Theory perspective to shorten tests show a psychometric cost that may limit the practical benefit. In this sense, within the IRT framework, an alternative solution is offered to optimize the evaluation of Neuroticism from Computerized Adaptive Testing.

Empirical evidence was obtained to corroborate the proposed hypotheses. The pool of 36 items that make up the bank constructed in the present research gathers acceptable psychometric properties for the valuation of individual differences in the Neuroticism domain

according to the FFM definitions. The items were previously submitted to the critique of expert judges and pilot tests were conducted in order to provide evidence of both content and apparent validity. Subsequently, the unidimensionality of the construct and the GRM item fit were corroborated (H1). The FIT revealed that the bank provides high information in a wide range of trait values while the correlation patterns with EPQ-R and SCL-90-R provided evidence of validity based on the relationship with other variables (H2). These results show that conventional bank administration has adequate psychometric features to justify efficiency analysis when applied in an adaptively format (H3).

One aspect that should be analyzed is the estimation of relatively lower *a* parameter values when calibrating Hostility and Impulsivity items. By definition, Neuroticism is a complex domain composed of a variety of negative feelings and emotions that are interrelated but also heterogeneous enough to conceptually isolate them into facet. In recent years some authors have pointed out that the inclusion of anger and impulse control as an essential part of Neuroticism entails a very risky theoretical compromise (e.g. Tackett & Lahey, 2017; Widiger, 2009). Both facets tend to stand out in validation studies because they present the lowest factor loadings in the Neuroticism factor and because they are associated with similar intensity to other model domains. Indeed, there are not few FFM theorists who proposed alternative taxonomies to that of McCrae & Costa (2010) and who have operationalized Neuroticism without including at least one of these two discussed facets (e.g. Aguado et al., 2008; Saucier, 2002; Soto & John, 2017b; Taylor & DeBruin, 2006; Watson et al., 2017). However, in the absence of a consolidated theoretical model that recognizes these variations in the delimitation of the construct, it was decided to maintain a top-down approach that could provide a basis for the bank's development.

In relation to the CAT methodology implemented here, it has been demonstrated that it is possible to obtain estimates of Neuroticism with an optimal degree of precision over much of the trait continuum, even when only part of the items that make up the bank are administered. For both of the CAT version the correlations of θ with the entire bank were high, surpassing even the most demanding criterion of $r \geq .95$ suggested by Thompson (2009). In the two CAT versions the standard errors of estimation at θ were low (equivalent to a classical minimum reliability of .90) even for those subjects who were two standard deviations above or below the trait's mean. When a stopping rule of fixed-length is considered only 50% of total bank's items were administered, while with variable-length stopping rule (error ≤ .25) this amount was reduced, on average, by 35%.

If the results of the CAT stopping rules are compared, it can be concluded that the 18-item-fixed-length variant was more efficient. Although it was necessary to administer more items, the mean of standard error of estimation = 0.22 was lower than the forecasted in the variable-length version (0.25). In addition, the fixed-length CAT estimates of θ were more strongly associated with both the θ of the whole bank and with the total score calculated based on the classical theory. These latter correlations are particularly important if they are interpreted taking into account the validity of the CAT content. The adaptive algorithm selects one item among all the available items based on a quality psychometric criterion (maximum information) regardless of the content of the item. As a consequence, in the variable-length CAT the under-representation of Hostility and Impulsivity items is worsened. The correlation registered between the θ estimated with the fixed-length CAT and the θ obtained with the complete bank was higher, it is observed that the content sampling had less impact on the construct measurement than variable-length CAT.

The saved administration time for Neuroticism measurement that comes as a result from applying an adaptive version with 18 items is considerable if the 48 items of the NEO-PI-3 or the 60 elements of the NEO-IPIP (Goldberg, et al., 2006) are taken as a reference. However, it is more extensive if it is compared with other instruments such as the NEO-FFI-3 (McCrae & Costa, 2010) or the BFI-2 (Soto & John, 2017b) which use 12 items.

In this regard, it is convenient to highlight two aspects that differentiate these conventional abbreviated tests from the measurements obtained with the CAT:

1) In the conventional short versions the coverage of the construct is usually reduced by eliminating the items with contents that do not show high discrimination capacity for the average feature. For example, in the NEO-FFI-3, Impulsivity is considered irrelevant for the brief Neuroticism measurement, so it does not include items that operationalize this content. On the other hand, although with different exposure rates, all items of the bank were used for some of the subjects adaptively evaluated in this study. This implies that the reduction in the number of items administered with the adaptive version does not restrict the representation of the facets in the trait measurement. The content is available in the bank and serves for the evaluation of individuals as long as it can provide information about the trait. Nonetheless, the current limitation of the bank's control of content coverage will be discussed further below.

2) The adaptive procedure offers guarantees of a measurement with a higher level of accuracy even for the extreme trait values, for which conventional tests are more error-prone (Aguado et al., 2005; Reise & Revicki, 2015). Given that it is precisely the extremely high Neuroticism scores that may be most relevant in the clinical context, it seems justified to measurably increase the number of items that would be applied with a short instrument in order to achieve a better precision in the trait estimation.

Conclusion

The results obtained in this study are encouraging because they show that the bank's current psychometric properties would enable an accurate and faster adaptive Neuroticism measurement than instruments using conventional forms of administration. However, one of the limitations of the adaptive design presented is the low control over the item's content that was effectively answered by each of the evaluatees. While content coverage in the entire bank has been guaranteed, the same is not true for the adaptive versions that the subjects have responded to. The strong correlations found between the θ estimates with the complete bank and with each CAT demonstrate that these variations in the selected content sampling for every evaluatee did not significantly affect the measurement of the construct. Even so, the programming of adaptive algorithms that regulate the representativeness of the facets CAT will be analyzed in future studies. To this end, the incorporation of new items into the bank is essential. A greater degree of specificity will be required to identify item contents applicable in the local culture that show more discriminatory capacity, especially for the Hostility and Impulsivity facets. The inclusion of these new items will allow to propose improvements in the adaptive procedure to further optimize the Neuroticism measurement.

Another line of complementary research that is being developed aims to carry out adaptive Neuroticism measurements at the facets level (Abal, et al., 2019). The debate about the convenience of assessing personality traits with narrow or broad measures has no conclusive answers (e.g. Ashton et al., 2014; Salgado, et al., 2015). But to obtain a measure of each facet would allow to reach a greater completeness in the description and prediction of the profiles of those evaluated. In this line, it will be the bank user who will decide whether to measure the domain or facets according to their evaluation objectives in the future.

Limitations of the study

At this stage of the Neuroticism bank construction, no differential functioning studies of the items (DIF) have been carried out, which constitutes a methodological limitation to the present study. The DIF analysis provides validity evidence that makes it possible to

guarantee that the bank's measurements are not conditioned by the belonging of an individual to a specific group. This would allow the detection of potential bias based on, for example, gender, age range, or even the clinical/non-clinical condition of the evaluated person.

Implications for future research

Further research will also seek to validate a cut-off on the Neuroticism scale in order to differentiate subjects with clinically significant levels. The study of an interruption modality based on clinical criteria can optimize adaptive measurement if it is intended to be used for evaluation tasks with screening purposes (Fonseca-Pedrero et al., 2013, Rudick et al., 2013). In this circumstance, administration may be briefer, because items that showed maximum discrimination capacity at the level of the trait associated with the cut-off point, are only selected.

Funding/Financial Support

University of Buenos Aires (UBACyT2018 20020170100200BA and UBACyT2020 20020190200156BA); National Agency of Scientific and Technological Promotion (ANPCyT, PICT-2017-3226).

Other Support/Acknowledgement

The authors have no support to report.

Competing Interests

The authors have declared that no competing interests exist.

References

Abal, F. J. P., Auné, S. E. & Attorresi, H. F. (2018). Variación de la escala Likert en el test de utilidad de la matemática. [Variation in Likert scale in the mathematics usefulness test]. *Interacciones, 4* (3). https://doi.org/10.24016/2018.v4n3.134

Abal, F. J. P., Auné, S. E. & Attorresi, H. F. (2019). Construcción de un banco de ítems de Facetas de Neuroticismo para el desarrollo de un test adaptativo [Constructing a bank of neuroticism facet items for the developmentof an adaptive test]. *Psicodebate, 19* (1), 31 -50. https://doi.org/10.18682/pd.v19i1.854

Abal, F. J. P., Lozzia, G. S., Aguerri, M. E., Galibert, M. S. & Attorresi, H. F. (2010). La escasa aplicación de la Teoría de Respuesta al Ítem en Tests de Ejecución Típica. [The limited application of the Item Response Theory in typical performance tests]. *Revista Colombiana de Psicología, 19* (1) 111-122.

Aguado, D., Rubio, V. J., Hontangas, P. M. & Hernández, J. M. (2005). Propiedades psicométricas de un test adaptativo informatizado para la medición del ajuste emocional. [Psychometric properties of an emotional adjustment computerized adaptive test]. *Psicothema, 17*, 484-491.

Aguado, D.; Lucia, B., Ponte, G., & Arranz, V. (2008). Análisis inicial de las Propiedades Psicométricas del Cuestionario BFCP Internet para la Evaluación de Big Five. [Initial analysis of the psychometric properties of the BFCP internet questionnaire for Big Five assessment] *Revista Electrónica de Metodología Aplicada, 13* (2), 15-30. doi: 10.17811/rema.13.2.2008.1530.

Ashton, M. C., Paunonen, S. V., & Lee, K. (2014). On the validity of narrow and broad personality traits: A response to Salgado, Moscoso, and Berges (2013). *Personality and Individual Differences, 56*, 24-28. https://doi.org/10.1016/j.paid.2013.08.019

Attorresi, H., Lozzia, G., Abal, F., Galibert, M. & Aguerri, M. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. [Item Response Theory. Basic concepts and applications for the measurement of psychological constructs] *Revista Argentina de Clínica Psicológica, XVIII,* 2, 179 - 188.

Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric Properties of the Mini-IPIP in a Large, Nationally Representative Sample of Young Adults. *Journal of Personality Assessment*, *95*(1), 74–84. https://doi.org/10.1080/00223891.2012.700466

Barlow, D. H., Ellard, K. K., Sauer-Zavala, S., Bullis, J. R. & Carl, J. R. (2014). The Origins of Neuroticism. *Perspectives on Psychological Science, 9* (5) 481–496. https://doi.org/10.1177/1745691614544528

Beiser, D., Vu, M., & Gibbons, R. (2016). Test-Retest Reliability of a Computerized Adaptive Depression Screener. *Psychiatric Services in Advance, 67* (9), 1039–1041. https://doi.org/10.1176/appi.ps.201500304

Byrne, B.M. (2012). *Structural equation modeling with Mplus: Basics, concepts, applications, and programming*. New York: Routledge.

Casullo, M. (2004). Síntomas psicopatológicos en adultos urbanos. [Psychopathological symptoms in urban adults.] *Psicología y Ciencia Social, 6* (1), 49-57.

Casullo, M.M. & Perez, M. (2008). *El Listado de Síntomas SCL-90-R de Derogatis.* [The Derogatis Symptom ListSCL-90-R]. Publications Department. Faculty of Psychology. University of Buenos Aires.

Chapman, B. P. & Goldberg, L. R. (2017). Act-Frequency Signatures of the Big Five. *Personality and Individual Differences, 116,* 201 - 205. https://doi.org/10.1016/j.paid.2017.04.049

Choi, S.W. (2009). Firestar: Computerized Adaptive Testing Simulation Program for Polytomous Item Response Theory Models. *Applied Psychological Measurement, 33* (8), 644-645. doi: 10.1177/0146621608329892.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, *102*(4), 874–888. https://doi.org/10.1037/a0027403

Cupani, M. & Lorenzo-Seva, U. (2016). The development of an alternative IPIP inventory measuring the Big-Five factor markers in an Argentine sample. *Personality and Individual Differences, 91*, 40–46. https://doi.org/10.1016/j.paid.2015.11.051

Cupani, M., Pilatti, A., Urrizaga, A., Chincolla, A. & Richaud, M.C. (2014). Inventario de personalidad IPIP-NEO: estudios preliminares de adaptación al español en estudiantes argentinos. [IPIP-NEO personality inventory: Preliminary studies of adaptation to Spanish in a sample of Argentinean students] *Revista Mexicana de Investigación en Psicología*, 6 (1), 55-73.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 Aspects of the Big Five. *Journal of Personality and Social Psychology, 93* (5), 880-896. https://doi.org/10.1037/0022-3514.93.5.880

Derogatis, L. (1994). *SCL-90-R. Symptom Checklist-90-R. Administration, Scoring and Procedures Manual*. National Computer System.

Devine, J., Fliege, H., Kocalevent, R., Mierke, A., Klapp, B. F. & Rose, M. (2016). Evaluation of Computerized Adaptive Tests (CATs) for longitudinal monitoring of depression, anxiety, and stress reactions. *Journal of Affective Disorders, 190*, 846–853. https://doi.org/10.1016/j.jad.2014.10.063

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment, 18* (2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Drake, M. M., Morris, D. & Davis, T. J. (2017). Neuroticism's susceptibility to distress: Moderated with mindfulness. *Personality and Individual Differences, 106,* 248-252. https://doi.org/10.1016/j.paid.2016.10.060

Drasgow, F., Levine, M. V., Tsien, S., Williams B.A., & Mead, A.D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19* (2), 143-165. https://doi.org/10.1177/014662169501900203

Eysenck, H.J., & Eysenck, S.B.G. (1994). *Manual of the Eysenck Personality Questionnaire*. California: EdITS/Educational and Industrial Testing Service.

Fonseca-Pedrero, E., Menéndez, F.L., Paino, M., Lemos-Giráldez, S., & Muñiz, J. (2013). Development of a computerized adaptive test for schizotypy assessment. *PLoS ONE 8* (9), e73201. https://doi.org/10.1371/journal.pone.0073201

Frølund Pedersen, H., Frostholm, L., Søndergaard Jensen, J., Ørnbøl, E., & Schröder, A. (2016). Neuroticism and maladaptive coping in patients with functional somatic syndromes. *British Journal of Health Psychology, 21* (4), 917-936. https://doi.org/10.1111/bjhp.12206

Gibbons, R. D., Weiss, D.J., Frank, E. & Kupfer, D. (2016). Computerized Adaptive Diagnosis and Testing of Mental Health Disorders. *Annual Review of Clinical Psychology, 12*, 83 - 104. https://doi.org/10.1146/annurev-clinpsy-021815-093634

Goldberg, L.R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4* (1), 26-42. https://doi.org/10.1037/1040-3590.4.1.26

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48* (1), 26-34. https://doi.org/10.1037/0003-066X.48.1.26

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., & Gough, H.C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Hajek, A; Bock, J.O. & König, H.H. (2017). The role of personality in health care use: Results of a population-based longitudinal study in Germany. *PLoS One, 12* (7):e0181716. https://doi.org/10.1371/journal.pone.0181716

Hengartner, M. P., Kawohl, W., Haker, H., Rössler, W., & Ajdacic-Gross, V. (2016). Big Five personality traits may inform public health policy and preventive medicine: Evidence from a cross-sectional and a prospective longitudinal epidemiologic study in a Swiss community. *Journal of Psychosomatic Research, 84*, 44 - 51. doi: 10.1016/j.jpsychores.2016.03.012.

Jeronimus, B. F., Kotov, R., Riese, H. & Ormel, J. (2016). Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly decays with time: a meta-analysis on 59 longitudinal/prospective studies with 443313 participants. *Psychological Medicine, 46* (14), 2883-2906. https://doi.org/10.1017/S0033291716001653

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist, 64*, 241–256. https://psycnet.apa.org/doi/10.1037/a0015309

Lozano, L.M., García-Cueto, E. & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology, 4* (2), 73–79. https://doi.org/10.1027/1614-2241.4.2.73

McCrae, R. R. & Costa, P. T. (2007). Brief Versions of the NEO-PI-3. *Journal of Individual Differences, 28* (3), 116 – 128. https://doi.org/10.1027/1614-0001.28.3.116

McCrae, R. R. & Costa P. T. (2010). *NEO Inventories professional manual.* Odessa, FL: Psychological Assessment Resources.

Milojev, P., Osborne, D., Greaves, L.M., Barlow, F.K. & Sibley, C.G. (2013). The Mini-IPIP6: Tiny yet highly stable markers of Big Six personality. *Journal of Research in Personality, 47*, 936–944. https://doi.org/10.1016/j.jrp.2013.09.004

Morizot, J. (2014). Construct validity of adolescents' self-reported Big Five personality traits: Importance of conceptual breadth and initial validation of a short measure. *Assessment, 21* (5), 580-606. https://doi.org/10.1177%2F1073191114524015

Muthén, L. & Muthén, B. (2010). *Mplus User's Guide, 6th Edn*. CA: Muthén & Muthén.

Natividade, J. C. & Hutz, C. S. (2015). Escala Reduzida de Descritores dos Cinco Grandes Fatores de Personalidade: Prós e Contras. [Short form scale of descriptors of the Big Five Personality Factors: Pros and Cons]. *Psico, 46* (1), 79-89. https://doi.org/10.15448/1980-8623.2015.1.16901

Nieto, M.D., Abad, F.J., Hernández-Camacho, A., Garrido, L.E., Barrada, J.R., Aguado, D. & Olea, J. (2017). Calibrating a new item pool to adaptively assess the Big Five. *Psicothema, 29* (3), 390-395. doi: 10.7334/psicothema2016.391

Reise, S. P. & Revicki, D. A. (Eds.). (2015). *Handbook of item response theory modeling applications to typical performance assessment*. Routledge/Taylor & Francis Group.

Reise, S.P. & Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: some empirical and conceptual issues and challenges. *Psychol Med, 46* (10), 2025-2039. https://doi.org/10.1017/S0033291716000520

Rubio, V. J., Aguado, D., Hontangas, P. M. & Hernández, J. M. (2007). Psychometric properties of an Emotional Adjustment Measure. An application of the Graded Response Model. *European Journal of Psychological Assessment, 23* (1), 39-46. https://doi.org/10.1027/1015-5759.23.1.39

Rudick, M. M., Yam, W. H., & Simms, L. J. (2013). Comparing countdown- and IRT-based approaches to computerized adaptive personality testing. *Psychol Assess, 25* (3), 769-79. https://psycnet.apa.org/doi/10.1037/a0032541

Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B., & Berges, A. (2015). Validity of the five-factor model and their facets: The impact of performance measure and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology, 24*, 325–349. https://psycnet.apa.org/doi/10.1080/1359432X.2014.903241

Samejima, F. (2010). Graded Response Model. En W. J. van der Linden (Ed.). *Handbook of Item Response Theory, Volume 1: Models* (pp. 95-108). Chapman y Hall/CRC.

Sánchez, R.O. & Ledesma, R.D. (2009). Análisis psicométrico del Inventario de Síntomas Revisado (SCL-90-r) en población clínica. [Psychometric analysis of the Revised Symptom Inventory (SCL-90-r) in clinical population]. *Revista Argentina de Clínica Psicológica, XVIII*, 265-274.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the big five. *Journal of Research in Personality*, 36 (1), 1 – 31. https://doi.org/10.1006/jrpe.2001.2335

Sauer-Zavala, S., Wilner, J. & Barlow, D. H. (2017). Addressing neuroticism in psychological treatment. *Personality Disorders: Theory, Research, and Treatment, 8* (3), 191-198. doi: 10.1037/per0000224.

Sibley, C. G. (2012). The Mini-IPIP6: Item Response Theory analysis of a short measure of the big-six factors of personality in New Zealand. *New Zealand Journal of Psychology, 41* (3), 21-31.

Soto, C. J. & John. O. P. (2017a). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81. https://psycnet.apa.org/doi/10.1016/j.jrp.2017.02.004

Soto, C. J., & John, O. P. (2017b). The Next Big Five Inventory (BFI2): Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *Journal of Personality and Social Psychology, 110* (3), 127. https://doi.org/10.1037/pspp0000096

Squillace, M., Picón Janeiro, J. & Schmidt, V. (2013). Adaptación local del Cuestionario Revisado de Personalidad de Eysenck (Versión abreviada). [Local adaptation of Eysenck Personality Questionnaire (short version)]. *Evaluar, 13*, 19 – 37. https://doi.org/10.35670/1667-4545.v13.n1.6794

Stark, S. (2007). *MODFIT: Plot theoretical item response functions and examine the fit of dichotomous or polytomous IRT models to response data* [computer program]. Champaign, IL: University of Illinois at Urbana-Champaign.

Stark, S., Chernyshenko, O.S., Drasgow, F., & White, L.A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods, 15*, 463-487. https://doi.org/10.1177%2F1094428112444611

Tackett, J.L. & Lahey, B.B. (2017). Neuroticism. En T. A. Widiger (Ed). *The Oxford handbook of the five factor model*. New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199352487.013.14.

Taylor, N., & De Bruin, G. (2013). The Basic Traits Inventory. In Laher S. & Cockcroft K. (Eds.), *Psychological Assessment in South Africa: Research and applications*. 232-243. Wits University Press. doi:10.18772/22013015782.21

ten Have, M., Oldehinkel, A., Vollebergh, W. & Ormel, J. *et al.* (2005). Does neuroticism explain variations in care service use for mental health problems in the general population? Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol*, *40* (6), 425-431. https://doi.org/10.1007/s00127-005-0916-z

Thissen, D. (2003). *MULTILOG*. Scientific Software International.

Thompson, N. (2009). *Ability estimation with item response theory.* Assessment Systems Corporation.

Vittengl, J. R. (2017). Who pays the price for high neuroticism? Moderators of longitudinal risks for depression and anxiety. *Psychological Medicine*, *47* (10) 1794-1805. https://doi.org/10.1017/S0033291717000253

Watson, D., Nus, E., & Wu, K. D. (2017). Development and validation of the faceted inventory of the Five-Factor Model (FI-FFM). *Assessment, 26* (1) 17–44. https://doi.org/10.1177/1073191117711022

Widiger, T. A. & Oltmanns, J. R. (2017). Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry, 16* (2), 144–145. https://doi.org/10.1002/wps.20411

Widiger, T. A. (2009). Neuroticism. En M. R. Leary & R. H. Hoyle (Eds), *Handbook of individual differences in social behavior* (p. 129-146). Guilford Press.

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*(4), 185-189. https://psycnet.apa.org/doi/10.1027/1614-0001/a000148

Appendix

Original and translated items of Bank.

1. Suelo tener contracturas o tensión muscular provocada por los nervios. // I usually suffer from muscle contractures or tension caused by nerves.

2. He dejado de hacer muchas actividades porque no me animé a asumir el riesgo. // I have stopped many activities because I didn't dare to take the risk.

3. Tengo miedo a muchas cosas. // I am afraid of many things.

4. A veces me doy cuenta de que estoy pensando muy rápido y no puedo frenar las ideas. // Sometimes I realize that I am thinking too fast and cannot stop.

5. A veces siento que me invento problemas en donde otros podrían actuar sin preocuparse. // Sometimes I feel like I make up problems in which others could react without worrying.

6. Me cuesta olvidar las cosas desagradables que me pasaron, vuelvo a recordarlas una y otra vez. // It is difficult for me to forget the unpleasant things that have happened to me, I remember them again and again.

7. Soy tan orgulloso que me cuesta aceptar cuando me equivoco. // I am so proud that I find it hard to accept when I am wrong.

8. Me enojo fácilmente. // I get angry easily.

9. Me fastidia que alguien me venga a molestar cuando estoy concentrado haciendo algo. // I am annoyed by someone coming to bother me when I am focusing on doing something.

10 (-). Puedo escuchar una crítica con calma. // I can hear criticism calmly.

11. Me cuesta perdonar a las personas, incluso cuando me piden disculpas por haberme ofendido. // I find it hard to forgive people even when they apologize for having offended me.

12. Me molesta cuando veo a personas que se pueden comprar todo lo que quieren. // It bothers me when I see people who can buy everything they want.

13. En algunas circunstancias me siento un inútil. // In some circumstances I feel useless.

14. A menudo me siento triste. // I often feel blue.

15. Siento que mi vida carece de sentido/dirección. // I feel that my life lacks direction.

16. Cuando las cosas salen mal suelo pensar que es por mi culpa. // When things go wrong I often think it's my fault.

17. No puedo evitar ver el aspecto negativo a las cosas que me pasan. // I can't help but see the negative side of things that happen to me.

18 (-). Mi estado de ánimo es bastante estable. // My mood is quite stable.

19. Me pongo muy incómodo en situaciones en las que debo ser el centro de atención de otras personas. // I get very uncomfortable in situations where I have to be the center of other people's attention.

20. Cuando estoy en grupo prefiero no hablar mucho para para evitar dar opiniones erradas. // When I am in a group I prefer not to talk much to avoid giving the wrong opinions.
21. Cuando las personas me miran en la calle supongo que están buscándome defectos. // When people look at me on the street I assume they are looking for my flaws.
22. Prefiero no asistir a reuniones en las que estoy seguro de que habrá personas a las que les caigo mal. // I prefer not to attend meetings where I am sure there will be people who dislike me.
23. Me resulta difícil acercarme a los demás. // I find it difficult to approach others.
24. Estoy pendiente de mi apariencia para evitar la crítica de las personas. // I keep an eye on my appearance to avoid criticism from other people.
25. A veces siento una necesidad incontrolable de comprar algo, aunque sea de poco valor. // Sometimes I feel an uncontrollable need to buy something, even if it's of little cheap.
26. Cuando quiero algo, lo quiero ya. // When I want something, I want it now.
27. No sé por qué realizo algunas cosas que cometo. // I don't know why I do some of the things I do.
28. En ocasiones suelo comer tanto que luego termino con alguna dolencia. // Sometimes I eat too much that I end up with some kind of affliction.
29. Suelo tomar decisiones precipitadas. // I tend to make rash decisions.
30. A veces siento deseos de romper cosas. // Sometimes I feel like smashing things.
31. Algunas veces me ha parecido que mis proyectos estaban tan llenos de dificultades que he tenido que abandonarlos. // Sometimes it has seemed to me that my projects were so full of difficulties that I had to abandon them.
32. Suelo quedarme paralizado en las situaciones de emergencia. // I tend to get paralyzed in emergency situations.
33 (-). No me dejo desalentar por los demás. // I do not let myself be discouraged by others.
34. Me siento incapaz de enfrentar las cosas. // I feel unable to face up to things.
35 (-). Tengo el talento suficiente como para superar con éxito todos los desafíos que se me presentan. // I am talented enough to successfully overcome all the challenges I face.
36. Me he encontrado con problemas tan llenos de alternativas que no he podido llegar a tomar una decisión. // I have faced problems so full of alternatives that I have not been able to make a decision.

About the Authors

Facundo Juan Pablo Abal works as a researcher in National Council of Scientific and Technological Investigations (CONICET). He has a PhD in Psychology (University of Buenos Aires, Argentina) and is a professor at the University of Buenos Aires.

Gabriela Susana Lozzia is an associate professor and research fellow at the Faculty of Psychology (University of Buenos Aires). She has a PhD in Psychology (University of Buenos Aires, Argentina).

Sofía Esmeralda Auné works as a researcher in National Council of Scientific and Technological Investigations (CONICET). She has a PhD in Psychology (University of Buenos Aires, Argentina) and is a professor at the University of Buenos Aires.

Horacio Félix Attorresi is an associate professor and research fellow at the Faculty of Psychology (University of Buenos Aires). He is director of several funded research projects on the Psychometrics.

Corresponding Author's Contact Address [TOP]

Facundo Juan Pablo Abal
Email: afjp79@hotmail.com