Almiro, P.A., Moura, O., & Simões, M.R. (2016). Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire – Revised (EPQ-R). *Personality and Individual Differences*, *88*, 88-93. doi:10.1016/j.paid.2015.08.050, ISSN: 0191-8869

Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire – Revised (EPQ-R)

Pedro Armelim Almiro[1]
Cognitive and Behavioural Center for Research and Intervention (CINEICC). Faculty of Psychology and Educational Sciences, University of Coimbra (Portugal).

Octávio Moura
CINEICC. Faculty of Psychology and Educational Sciences, University of Coimbra (Portugal).

Mário R. Simões
Laboratory of Psychological Assessment and Psychometrics. CINEICC.
Faculty of Psychology and Educational Sciences, University of Coimbra (Portugal).

[1] Corresponding author. E-mail address: psi.armelim22@yahoo.com (P.A. Almiro).

Abstract

Based on the P-E-N Model, the Eysenck Personality Questionnaire – Revised (EPQ-R; S. Eysenck et al., 1985) is an internationally well-known personality assessment instrument. This questionnaire measures the three fundamental personality dimensions: Psychoticism, Extraversion, and Neuroticism (also includes a Lie/Social Desirability scale).

The aim of the present paper consists in the examination of the factorial structure of the EPQ-R in the Portuguese context and its psychometric properties (validity and reliability). Using a large sample (*N*=1689, 16-60 years), the construct validity was examined through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), and the test reliability through internal consistency analysis and test-retest reliability (temporal stability between 4 to 8 weeks).

In general, the EPQ-R showed an adequate reliability and validity indices, replicating on the Portuguese context the factor structure of its original version (English). These results demonstrate that the EPQ-R conveniently measures the Neuroticism, Extraversion, and Psychoticism constructs, defined by H. Eysenck, and the adequacy of the EPQ-R to the Portuguese population as a personality assessment instrument.

Keywords: Eysenck Personality Questionnaire – Revised (EPQ-R), exploratory factor analysis, confirmatory factor analysis, P-E-N Model, personality assessment.

1. Introduction

According to H. Eysenck (1970, p.2), personality can be defined as «the more or less stable and enduring organization of a person's character, temperament, intellect, and physique, which determines his unique adjustment to the environment»; character denotes a person's more or less stable and enduring system of conative behaviour (will); temperament, the system of affective behaviour (emotion); intellect, the system of cognitive behaviour (intelligence); physique, the system of bodily configuration and neuro-endocrine endowment.

The Eysenck Personality Questionnaire – Revised (EPQ-R; S. Eysenck, H. Eysenck, & Barrett, 1985; European Portuguese version, Almiro & M. Simões, 2013) is an internationally well-known personality assessment instrument, which has been used in several application contexts (e.g., clinical, forensic, health, educational, organizational, military) (see Lynn, 1981; Nyborg, 1997). There are several psychometric studies of this instrument, which were carried out in more than thirty-four countries, including different cultures over the five continents (e.g., Europe: England, Germany, Italy, Spain; America: Brazil, Canada, United States; Asia: India, Japan; Oceania: Australia; Africa: Egypt, Uganda) (see Barrett, Petrides, S. Eysenck, & H. Eysenck, 1998; S. Eysenck & Barrett, 2013; Furnham, S. Eysenck, & Saklofske, 2008).

The EPQ-R is a self-report questionnaire constructed to measure the three-dimensional personality taxonomy proposed by H. Eysenck in the P-E-N Model: *Psychoticism* (P), *Extraversion* (E), and *Neuroticism* (N). In the Eysenck's perspective, these are the three fundamental personality dimensions (*Big Three*) to describe the emotional and behavioural human characteristics (character and temperament) of the most importance and they are assessed through the four scales of EPQ-R: P, E, N scales, and additionally the L scale to measure the Lie/Social Desirability construct, which constitutes a validity scale (H. Eysenck & S. Eysenck, 2008; Furnham et al., 2008).

The N dimension assesses the continuum between two idealized extremes: on one hand, the neurotic or emotional unstable personality, and on the other hand, the emotional stable personality. The typical neurotic subject is anxious, depressed, emotional, tense, shy, moody, worried, is likely to sleep bad, has guilty feelings, has low self-esteem, suffers from various psychosomatic disorders, and reacts too strongly to all sorts of stimuli. The typical stable subject has the opposite traits and he is calm, controlled, steadfast, easygoing, and even-tempered.

The E dimension also measures a continuum between two extremes: the extrovert personality and the introvert personality. The typical extrovert is sociable, lively, active, talkative, assertive, dominant, surgent, outgoing, carefree, optimistic, venturesome, sensation-seeker, likes changes, has a ready answer, takes chances and acts in the spur of the moment. The typical introvert has the opposite traits and he is unsociable, quiet, introspective, passive, thoughtful, reserved, pessimistic, sober, reliable, and peaceful.

The P dimension, like the others, measures a continuum between two extremes: the psychoticism (tough-mindedness) and the adjusted personality (impulse control). The subject with high psychoticism is aggressive, impulsive, cold, impersonal, insensitive, egocentric, solitary, suspicious, antisocial, non-empathic, tough-minded, troublesome, creative, he is hostile even to loved ones, likes to make fools of other people and to upset them, has a liking for odd and unusual things, and has a disregard for danger. These traits are the opposite of a subject who has an adjusted personality, being empathic, socialized, altruistic, tolerant, conventional, conformist, responsible, conscientious, friendly, agreeable, and warm (H. Eysenck, 1994; H. Eysenck & M. Eysenck, 1985; H. Eysenck & S. Eysenck, 1975).

The EPQ-R was preceded by the development of successive personality questionnaires for over the past fifty years of investigation: Maudsley Medical Questionnaire (MMQ), Maudsley Personality Inventory (MPI), Eysenck Personality Inventory (EPI), Eysenck Personality Questionnaire (EPQ), and EPQ-R (Dazzi, 2011; S. Eysenck & Barrett, 2013; Furnham et al., 2008). The EPQ-R is the revised version of the EPQ, which was developed to improve the psychometric limitations of P scale (low reliability and low range of scoring). Its factor structure was tested through the exploratory factor analysis (principal component analysis; varimax, promax and direct oblimin). In its original version (English), the EPQ-R has 100 items distributed in four factors: P (32 items); E (23 items); N (24 items); and L (21 items). The internal consistency is: .78 for P, .90 for E, .88 for N, and .82 for L in the male sample; .76 for P, .85 for E, .85 for N, and .79 for L in the female sample (S. Eysenck et al., 1985). The EPQ had 90 items (four factors): P (25 items); E (21 items); N (23 items); and L (21 items). The internal consistency was: .74 for P, .85 for E, .84 for N, and .81 for L in the male sample; .68 for P, .84 for E, .85 for N, and .79 for L in the female sample (H. Eysenck & S. Eysenck, 1975). In the European Portuguese version of the EPQ (73 items), studied in 1991, the results were quite similar (except for P scale): .75 for P, .84 for E, .81 for N, and .83 for L in the male sample; .59 for P, .84 for E, .85 for N, and .81 for L in the female sample (Fonseca, S. Eysenck, & A. Simões, 1991).

The aim of the present study is to examine the factorial structure of the EPQ-R in the Portuguese context and its psychometric properties in terms of reliability and construct validity.

2. Method

2.1. Participants

Participants are 1689 volunteers, 783 males (46.36%) and 906 females (53.64%), aged between 16 and 60 years old (age groups: 16-20, 21-30, 31-40, 41-50, 51-60), belonging to different professional categories and Portuguese regions. The mean age is 32.58 (SD=11.66) for males and 32.13 (SD=10.81) for females. The sample is wide and representative of the Portuguese population. It was considered the population distribution in proportion (Portugal: North, 35.76%, n=604; Center, 22.02%, n=372; Lisbon, 25.40%, n=429; Alentejo, 8.00%, n=135; Algarve, 3.97%, n=67; Azores, 2.78%, n=47; Madeira, 2.07%, n=35) and the geographic localization (coast, 80.70%, n=1363; inland, 19.30%, n=326) variables, based on the data from Instituto Nacional de Estatística (2012). Participants were recruited from various community contexts (e.g., schools, companies, associations, among others), using the nonprobability sampling method (convenience sampling and in some cases using a snowball sampling strategy). Participants were asked for voluntary participation and the objectives and relevance of the present study were explained to them (no incentives were offered in exchange for participation). Then, the informed consent information was gathered. Participants were asked to answer the questionnaire individually or in group, informing that their responses would remain anonymous and confidential.

2.2. Instrument

The European Portuguese version of the Eysenck Personality Questionnaire – Revised (EPQ-R; Almiro & M. Simões, 2013) is a 70 items self-report questionnaire used to assess three dimensions of personality: N (23 items); E (20 items); P (9 items); and L scale (18 items). The response to each item is “yes” or “no” (dichotomic) and the quotation for some items is inverted. The items of the EPQ-R were translated from its English (S. Eysenck et al., 1985), Spanish (H. Eysenck & S. Eysenck, 2008), and German (Ruch, 1999) versions to Portuguese language and then back-translated. Some items from the Portuguese version of the EPQ (Fonseca et al., 1991) were also used in the construction of the EPQ-R (Almiro & M. Simões, 2013).

2.3. Statistical analyses

To examine the construct validity of the European Portuguese version of the EPQ-R (Almiro & M. Simões, 2013), an *exploratory factor analysis* (EFA) followed by a *confirmatory factor analysis* (CFA) were performed using the SPSS (version 17.0) and the EQS (version 6.1; Bentler, 2006), respectively. The *correlations between factors*, the *item-factor correlations*, and the *corrected item-factor correlations* (discrimination index) were also performed (Pearson’s *r*). The reliability of the EPQ-R was examined through the *internal consistency* (Cronbach’s alpha) and the *test-retest reliability* methods for each scale (N, E, P, L) using the SPSS.

3. Results

3.1. Descriptive statistics and mean differences

Table 1 presents the mean and standard deviations obtained for each dimension of the EPQ-R for the whole sample (N=1689), males (n=783) and females (n=906). The mean differences between males and females were examined using the independent-samples *t-test*.

Table 1

These results showed that the means were significantly different (p<.001) between males and females for N [t(1669.476)=-7.630; d=-0.37], P [t(1687)=7.183; d=0.35], and L [t(1635.192)=-4.782; d=-0.24] scales. For E scale there were no significant differences. The corresponding Cohen's d indices showed a medium effect size for these t-values (cf. Cohen, 1988). Therefore, females tend to obtain higher scores in the N and L scales than males, and males tend to obtain higher scores in the P scale. These gender differences can be found in the original English version of the EPQ-R (S. Eysenck et al., 1985), and in the European Portuguese version of the EPQ (Fonseca et al., 1991).

3.2. Reliability

The alpha coefficients obtained in this study for the whole sample (N=1689) were the following: .87 for N, .83 for E, .55 for P, and .78 for L (see Table 2). These coefficients ranged from good to satisfactory, except for the P scale, which was below the minimum value of .70 pointed out by several authors (e.g., P. Kline, 1993; Nunnally & Bernstein, 1994). According to the criterion established by DeVellis (1991), these coefficients were considered "very good" for N and E, "respectable" for L, and "unacceptable" for P.

Table 2

In spite of these data, test-retest reliability analysis (*N*=124, independent sample; temporal stability between 4 to 8 weeks) yielded a contrasting coefficient for P, a better index for L, and maintained good coefficients for N and E scales. All coefficients ranged from good to satisfactory, assuming values above .70: .86 for N, .89 for E, .72 for P, and .86 for L (see Table 2). According to the same criterion (DeVellis, 1991), the reliability coefficients were considered "very good" for N, E, L, and "respectable" for P, which indicate that the results obtained through the EPQ-R assessment had good temporal stability.

3.3. Validity

The examination of the factorial structure of the Portuguese EPQ-R (construct validity) was performed applying EFA and CFA of the present data. In the EFA, several factor structures (three-, four- and five-factor solutions) were tested through the principal component analysis with direct oblimin and varimax methods (like the original study, see S. Eysenck et al., 1985) in order to get the most adjusted one. The four-factor solution, corresponding to the four dimensions of the instrument (N, E, P, L), showed a better adequacy and a good fit to the one obtained by H. Eysenck (cf. H. Eysenck & S. Eysenck, 1975; S. Eysenck et al., 1985). The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy was .88 and the Bartlett's test of sphericity was significant and yielded adequate values [χ^2=24433.60; df=2415; *p*=.000].

The procedure of factor extraction was simultaneously based on the Kaiser-Guttman's criterion (eigenvalues≥1.0) and the Cattell's scree test (cf. Nunnally & Bernstein, 1994). P. Kline (1994, p.52) refers that «a factor loading of .30 indicates that 9 per cent of the variance is accounted for by the factor; this is taken as large enough to indicate that the loading is salient; thus in factor analyses where the sample is at least 100 subjects this is a reasonable criterion; loadings of .30 or larger are regarded as significant». Thus, to retain the items with a significant loading it was considered the criterion proposed by P. Kline (1994).

Table 3

The four factors extracted were formed by 70 items and accounted for 25.94% of the total variance. The first factor, with 23 items, was defined by Neuroticism (N) and accounted for 10.39% of the variance, corresponding to an eigenvalue of 7.27. The factor loadings ranged between .69 and .38 (mean .50). The second factor, with 20 items, was the Extraversion (E) and accounted for 7.00% of the variance, corresponding to an eigenvalue of 4.90. The factor loadings ranged between .67 and .33 (mean .49). The third factor, with 18 items, was the Lie/Social Desirability (L) and accounted for 5.35% of the variance, corresponding to an eigenvalue of 3.75. The factor loadings ranged between .61 and .35 (mean .45). The fourth factor, with 9 items, was the Psychoticism (P) and accounted for 3.20% of the variance, corresponding to an eigenvalue of 2.24. The factor loadings ranged between .59 and .28 (mean .39). All significant factor loadings ranged between .69 and .31, except for the item 13 (“Would it upset you a lot to see a child or a animal suffer?”) which loading was .28 (a value close to .30 used as the criterion for the significant loadings). Using the varimax extraction method, the results were very similar to these ones (similar loadings and percentage of total variance explained).

All the correlations between the four factors (N, E, P, L) were significant (p<.01) and low (.30<r<.10) or null (r<.10) according to the criterion proposed by Cohen (1988), as expected. These coefficients ranged between -.20 and .23. In addition, the item-factor correlations (including the item) were significant (p<.01) and high (r>.50) or medium (.50<r<.30), ranging between: .67 and .39 (mean .51) for N; .65 and .34 (mean .49) for E; .57 and .30 (mean .46) for P; .58 and .39 (mean .46) for L. The corrected item-factor correlations (excluding the item) were satisfactory: .62 and .33 (mean .44) for N; .58 and .26 (mean .41) for E; .35 and .18 (mean .25) for P; .48 and .28 (mean .36) for L.

Based on these data, a CFA was performed over the variance-covariance matrix in order to test the replicability of the four-factor model (N, E, P, L) assessed by the EPQ-R. This model was estimated using the maximum likelihood method, following the procedures taken in the Dazzi (2011)’s psychometric study of the Italian version of this instrument. The goodness-of-fit indices considered were: chi-square (χ^2); ratio of chi-square to degrees of freedom (χ^2/df); comparative fit index (CFI); standardized root mean square residual (SRMR); and, root mean square error of approximation (RMSEA). A confirmatory model has a good fit when the ratio χ^2/df<3, CFI>.95, SRMR<.08, and RMSEA<.06 (R. Kline, 1998; Hu & Bentler, 1999). In spite of these widely-accepted criteria, the ratio χ^2/df<5 is also considered adequate if the CFA model in examination is a complex one (which was the case) (Schumacker & Lomax, 2004).

The present confirmatory model was tested with four latent variables, corresponding to N, E, P, and L factors of the EPQ-R. On a previous step, all four factors were tested separately though CFA in order to verify their unidimensionality. The results were satisfactory by showing the unidimensionality of each four factors and confirming the EFA results (good fit and significant loadings).

Like in Dazzi (2011)'s study, for the CFA we used the parceling of items through the item-to-construct balance method proposed by Little, Cunningham, Shahar, and Widaman (2002) in order to get a more optimal indicator to sample size ratio. According to Little et al. (2002)'s method, we used the factor loadings as a guide to construct parcels. For example, for the N factor the parcels were constructed by using the five items with the highest loadings to anchor the five parcels, the five items with the next highest item-to-construct loadings were added to the anchor in an inverted order (the highest loaded item from among the anchor items were matched with the lowest loaded item from among the second selections), and so on.

As Meade and Kroustalis (2006, p.371) pointed out, «adequate fit in the CFA model can be problematic for long surveys when items are used as indicators; however, the use of parcels can significantly improve model fit (…); parceling is a process by which raw item responses are combined into subscales prior to analysis; this is commonly done by summing or averaging item responses into parcel scores, which are then used as the lowest-order indicator variables in CFA». The use of parcels in CFA has many advantageous properties, including a greater reliability than individual items, a higher communality (larger ratio of common-to-unique variance), distributions with a more approximate to normality than with interval scale, a less item-idiosyncratic influence, a better model fit, and so on (cf. Bandalos & Finney, 2001; Little et al., 2002; Meade & Kroustalis, 2006).

Table 4

The 70 items of the Portuguese EPQ-R factor model were tested through 18 indicators originated by the aggregation of its items in parcels. Five parcels were produced for each of N, E, L factors and three parcels for P. Table 4 shows the established parcels for all four factors to produce the 18 indicators examined in the CFA (global model).

Figure 1

In this analysis, the four-factors were allowed to correlate with each other and there was no need to respecify the model including covariances between errors. Figure 1 shows the path diagram of the EPQ-R confirmatory factor analysis model tested (standardized solution), which the corresponding goodness-of-fit indices were: $\chi^2(129)=533.786$, $p<.001$; $\chi^2/df=4.13$; CFI=.961; SRMR=.042; RMSEA=.043. All factor loadings were significant and ranged between: .80 and .74 for N; .78 and .68 for E; .63 and .51 for P; .70 and .59 for L. Furthermore, as expected, the latent variables correlations weren’t significant and were low or null, ranging between -.29 and .31.

Additionally, we conducted a multiple-group analysis to find out whether the factor structure of the EPQ-R would be the same across males (n=783) and females (n=906). Following Byrne (2006) and Vandenberg and Lance (2000) suggestion, we tested measurement invariance based on the analysis of mean and covariance structures that encompassed a series of hierarchically ordered steps. So, it began with the establishment of a baseline model for each group (males and females), followed by tests for increasingly more stringent levels of constrained equivalence across both groups: configural, metric ("weak factorial invariance") and scalar invariance ("strong factorial invariance"). To determine evidence of invariance we compared the difference values of χ^2, df and CFI between the configural and the other two models (metric and scalar). Byrne and van de Vijver (2010) and Cheung and Rensvold (2002) recommend the following criteria for evidence of measurement invariance: (i) the multiple-group model exhibit an adequate fit to the data, (ii) the χ^2 difference value ($\Delta\chi^2$) is not statistically significant (p>.05), and (iii) the CFI difference value (ΔCFI) is <.010.

The EPQ-R factor model (baseline model) yielded an adequate model fit for males [$\chi^2(129)$=290.023, p<.001, χ^2/df=2.24, CFI=.968, SRMR=.043, RMSEA=.040] and females [$\chi^2(129)$=375.343, p<.001, χ^2/df=2.91, CFI=.954, SRMR=.045, RMSEA=.046]. After establishing the baseline for each group, we tested for configural model, wherein no equality constraints were imposed on the parameters across the two groups. The goodness-of-fit for the configural model was adequate, suggesting that both the number and pattern of factors were equivalent across groups (see Table 5). This allowed for the assessment of metric invariance by constraining all factor loadings to be equal across groups. Constraint of factor loadings across males and females resulted in a non-significant worse model fit according to the $\Delta\chi^2$=18.750, p =.175 and ΔCFI=.001, indicating that full metric invariance was met. Constraint of the factor loadings and intercepts resulted in a significant decrease of model fit [$\Delta\chi^2$=304.196, p<.001, ΔCFI=.026], suggesting that scalar invariance was not achieved. Successive examination of the probability values associated with the χ^2 univariate increment information provided by the Lagrange Multiplier Test for each parameter constraint revealed five invariant intercepts (E_P2, E_P4, E_P5, L_P2 and L_P5) and 13 non-invariant intercepts. If these 13 non-invariant intercepts were allowed to be freely estimated in each group the partial scalar invariance is supported [$\Delta\chi^2$=28.243, p=.079, ΔCFI=.001].

Table 5

4. Discussion

On the reliability analyses of the EPQ-R, all Cronbach's alphas ranged from "very good" to "respectable" (between .87 and .78), except for P which coefficient was "unacceptable" (.55). These values are similar to the original version (S. Eysenck et al., 1985), ranging between .90 and .76. Like the present Portuguese version, in the Italian EPQ-R (Dazzi, 2011) the alpha coefficients ranged between .85 and .75 for N, E, L, and was .67 for P. Some authors (e.g., Anastasi, 1988; Nunnally & Bernstein, 1994) refer that reliability indices can be influenced by the test/scale length. Shorter scales have some tendency to show lower coefficients than the longer ones. For example, the original P scale (S. Eysenck et al., 1985) is composed by 32 items, and the P scale of the Spanish version (H. Eysenck & S. Eysenck, 2008) has 23 items, whereas the P scale of the Portuguese EPQ-R only has 9 items. Because of the content in the P scale items, a plausible attitude of faking good in this community sample, caused by the social desirability, and the low range of scoring may had some influence on its results.

The results obtained through the test-retest reliability method (temporal stability) were quite good, ranging from "very good" to "respectable", between .89 and .72 (N=124; 4-8 weeks). These results are similar to the Spanish version (H. Eysenck & S.

Eysenck, 2008), with coefficients ranging between .86 and .72 (*N*=155; 4 weeks), and to the Italian version (Dazzi, 2011), with coefficients ranging between .93 and .79 (*N*=92; 4 weeks). Because of the contrasting indices obtained for P scale (Cronbach's alpha of .55, and test-retest reliability of .72), its reduced length (only 9 items) might had some real influence on these results. Except the Cronbach's alpha for P, all the reliability indices are greater than the minimum criterion of .70 consensually accepted (DeVellis, 1991), which means that the personality assessment of the EPQ-R produces reliable results and they have a good temporal stability.

On the validity examination of the EPQ-R, the results observed in the EFA and CFA demonstrated its adequacy to assess the Neuroticism, Extraversion, and Psychoticism constructs and the Lie/Social Desirability construct measured by L. Therefore, the applicability of the EPQ-R as a personality assessment instrument was also demonstrated with these results. The EFA clearly showed that a four-factor solution (N, E, P, L) were the most adequate to explain the data, which is a replication of the cross-cultural studies carried out in more than thirty-four countries (see Barrett et al., 1998; Furnham et al., 2008), and the CFA confirmed this finding through the adequacy of goodness-of-fit indices [$\chi^2(129)=533.786$, $p<.001$; $\chi^2/df=4.13$; CFI=.961; SRMR=.042; RMSEA=.043]. These results are similar to the model fit indices reported by Dazzi (2011). Furthermore, the test of measurement invariance (between males and females) conducted in the CFA showed the partial scalar invariance, suggesting that the same test score interpretation can be made across gender. Indeed, configural and full metric invariance was found, which means that the factor structure and the strength of the relationship between item parcels and their associated latent factors were equivalent across males and females. A more stringent level of constrained equivalence using scalar invariance yielded 13 non-invariant intercepts, indicating that these parcels means were not similar across gender when the latent trait value equals zero (males may obtain significantly different scores on these 13 parcels relative to females with an equivalent score in the latent factor). Partial scalar invariance was achieved after releasing constraints on the intercepts of these parcels.

In the EFA, the low total variance explained (ranging between 20% and 30%) is similar to other studies (e.g., Alexopoulos & Kalaitzidis, 2004; Aluja, Ó. Garcia, & L. García, 2003). The item 13 of the P scale was maintained in the factor structure because of its remaining psychometric properties: the alpha coefficient for P was higher when the item 13 was considered than when it wasn't; the item-factor correlation was .30 (positive), which is a medium coefficient according to the Cohen (1988)'s criterion; the corrected item-factor correlation was .18 (positive), which indicates a moderate discrimination index according to the criterion proposed by Nunnally and Bernstein (1994); the validity of P factor model, including the item 13, was examined through the CFA and showed an adequate fit.

Additionally, the correlations between factors (N, E, P, L) tended to be null (or low) and the correlations item-factor were positive and tended to be high (including and excluding the item), which indicate: a good convergence of each item to its respective factor (according to the original version); a good discrimination index of the items; and along with the EFA and CFA results, the unidimensionality and orthogonality of the factors (each factor only measures one major construct and the factors are relatively independent from each other). These EFA and CFA results are an extremely important evidence of the construct validity of the EPQ-R and its adequacy for use in the Portuguese context. However, the limitations related to the psychometric properties of the P scale – low internal consistency and low range of scoring – should be analyzed in subsequent studies with special groups, where the P traits are prominent: in the clinical context, with psychotic and bipolar patients, and in the forensic context, with psychopaths and criminal personalities.

In conclusion, the EPQ-R (70 items; Almiro & M. Simões, 2013) showed an adequate reliability and validity indices, replicating in the Portuguese context the factor structure of its original English version (S. Eysenck et al., 1985). These results demonstrate that the EPQ-R conveniently measures the Neuroticism, Extraversion, and Psychoticism constructs defined by H. Eysenck (cf. H. Eysenck & M. Eysenck, 1985; H. Eysenck & S. Eysenck, 1975) and confirm the universality of P-E-N Model. They also represent an important addition to the empirical support obtained for more than fifty years of research in personality theory and measurement carried out by H. Eysenck and his collaborators worldwide (cf. Barrett et al., 1998; S. Eysenck & Barrett, 2013; Furnham et al., 2008). In this sense, the EPQ-R can be a very useful instrument in the personality assessment, allowing to get a wide description of the subjects' behavioural and emotional characteristics related to their character, temperament, intellect, and physique aspects. These data can be analysed in terms of N, E and P dimensions, through a comprehensive system of traits. And L is a validity scale which assesses the person's proneness for lying or faking good (social desirability), therefore it can be very useful to examine the level of sincerity of the responses on the EPQ-R.

References

Alexopoulos, D.S., & Kalaitzidis, I. (2004). Psychometric properties of Eysenck Personality Questionnaire Revised (EPQ-R) Short scale in Greece. *Personality and Individual Differences*, *37*, 1205-1220.

Almiro, P.A., & Simões, M.R. (2013). *Manual da versão portuguesa do Questionário de Personalidade de Eysenck – Forma Revista (EPQ-R)* [Manual of the Portuguese version of the Eysenck Personality Questionnaire – Revised (EPQ-R)]. Coimbra: Laboratório de Avaliação Psicológica. In preparation.

Almiro, P.A., & Simões, M.R. (2014). Questionário de Personalidade de Eysenck – Forma Revista (EPQ-R) [Eysenck Personality Questionnaire – Revised (EPQ-R)]. In L.S. Almeida, M.R. Simões, & M.M. Gonçalves (Eds.), *Instrumentos e contextos de avaliação psicológica* – Vol. II (pp.211-229). Coimbra: Almedina.

Aluja, A., García, Ó., & García, L.F. (2003). A psychometric analysis of the Revised Eysenck Personality Questionnaire Short Scale. *Personality and Individual Differences*, *35*, 449-460.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: MacMillan.

Bandalos, D.L., & Finney, S.J. (2001). Item parceling issues in structural equation modeling. In G.A. Marcoulides (Ed.), *New developments and techniques in structural equation modelling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum Associates.

Barrett, P.T., Petrides, K.V., Eysenck, S.B., & Eysenck, H.J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences*, *26*, 805-819.

Byrne, B.M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, B.M., & van de Vijver, F. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*, 107-132.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dazzi, C. (2011). The Eysenck Personality Questionnaire – Revised (EPQ-R): A confirmation of the factorial structure in the Italian context. *Personality and Individual Differences*, *50*, 790-794.

DeVellis, R.F. (1991). *Scale development: Theory and applications*. London: SAGE.

Eysenck, H.J. (1970). *The structure of human personality* (3rd ed.). London: Methuen & Co.

Eysenck, H.J. (1994). The Big Five or Giant Three: Criteria for a paradigm. In C.F. Halverson, G.A. Kohnstamm, & R.P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp.37-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Eysenck, H.J. (1998). *Dimensions of personality*. New Brunswick: Transaction Publishers.

Eysenck, H.J., & Eysenck, M.W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum Press.

Eysenck, H.J., & Eysenck, S.B. (1975). *Manual of the Eysenck Personality Questionnaire (Junior & Adult)*. London: Hodder & Stoughton.

Eysenck, H.J., & Eysenck, S.B. (2008). *EPQ-R: Cuestionario Revisado de Personalidad de Eysenck: Versiones completa (EPQ-R) y abreviada (EPQ-RS)* [Revised Eysenck Personality Questionnaire: Complete (EPQ-R) and abbreviated (EPQ-RS) versions] (3a ed.). Madrid: TEA Ediciones.

Eysenck, S.B., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*, *54*, 485-489.

Eysenck, S.B., Eysenck, H.J., & Barrett, P.T. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences*, *6*, 21-29.

Fonseca, A., Eysenck, S.B., & Simões, A. (1991). Um estudo intercultural da personalidade: Comparação de adultos portugueses e ingleses no EPQ [A cross-cultural study of personality: A Comparison of Portuguese and English adults on the EPQ]. *Revista Portuguesa de Pedagogia*, *25*, 187-203.

Furnham, A., Eysenck, S.B., & Saklofske, D.H. (2008). The Eysenck personality measures: Fifty years of scale development. In G.J. Boyle, G. Matthews, & D.H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment – Vol. 2: Personality measurement and testing* (pp.199-218). London: SAGE.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Instituto Nacional de Estatística (2012). *Censos 2011: Resultados definitivos do XV Recenseamento Geral da População e do V Recenseamento Geral da Habitação* (8 Vols.) [Census 2011: Final results of the XV Census for Resident Population and of the V Census for Habitation]. Lisboa: Instituto Nacional de Estatística – Portugal.

Kline, P. (1993). *Personality: The psychometric view*. London: Routledge.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Kline, R.B. (1998). *Principles and practices of structural equation modeling*. New York: The Guilford Press.

Little, T.D., Cunningham, W.A., Shahar, G., & Widaman, K.F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151-173.

Lynn, R. (Ed.) (1981). *Dimensions of personality: Papers in honour of H.J. Eysenck*. Oxford: Pergamon Press.

Meade, A.W., & Kroustalis, C.M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods, 9*, 369-403.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Nyborg, H. (Ed.) (1997). *The scientific study of human nature: Tribute to Hans J. Eysenck at eighty*. Oxford: Pergamon Press.

Ruch, W. (1999). *Eine Untersuchung der revidierte Fassung des Eysenck Personality Questionnaire und die Konstruktion einer deutschen Standard- und Kurzfassung (EPQ-R bzw. EPQ-RK)* [An investigation of the Revised Eysenck Personality Questionnaire and construction of the German standard and short versions (EPQ-R and EPQ-RK)]. Düsseldorf: Heinrich-Heine-Universität Düsseldorf.

Schumacker, R.E., & Lomax, R.G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*, 4-70.

Table 1

Means, standard deviations, and differences in dimensions of the EPQ-R.

		N	E	P	L
total	M	10.44	12.61	1.01	9.68
	SD	5.53	4.43	1.31	3.72
males	M	9.35	12.66	1.26	9.22
	SD	5.31	4.64	1.46	3.76
females	M	11.37	12.57	0.81	10.08
	SD	5.55	4.24	1.12	3.64
t-test	*t*	-7.630*	0.429	7.183*	-4.782*
	sig.	.000	NS	.000	.000

Note. N = Neuroticism. E = Extraversion. P = Psychoticism. L = Lie/Social Desirability. M = mean. SD = standard deviation. * *t*-test value is significant; NS (*t*-test value is not significant).

Table 2

Reliability coefficients of the EPQ-R (70 items).

scales	internal consistency	test-retest reliability	items
N	.87	.86	23
E	.83	.89	20
P	.55	.72	9
L	.78	.86	18

Note. N = Neuroticism. E = Extraversion. P = Psychoticism. L = Lie/Social Desirability.

Table 3

Data from the EFA (principal component analysis, direct oblimin rotation).

	designation	items	loadings (range)	loadings (mean)	eigenvalues	% variance
1st factor	N	23	.69 – .38	.50	7.27	10.39%
2nd factor	E	20	.67 – .33	.49	4.90	7.00%
3rd factor	L	18	.61 – .35	.45	3.75	5.35%
4th factor	P	9	.59 – .28	.39	2.24	3.20%

Note. N = Neuroticism. E = Extraversion. P = Psychoticism. L = Lie/Social Desirability.

Table 4

Parcels for each factor and number of items on CFA.

factors	1st parcel	2nd parcel	3rd parcel	4th parcel	5th parcel
N (23 items)					
name of parcels	N_P1	N_P2	N_P3	N_P4	N_P5
number of items	4	4	5	5	5
E (20 itens)					
name of parcels	E_P1	E_P2	E_P3	E_P4	E_P5
number of items	4	4	4	4	4
P (9 itens)					
name of parcels	P_P1	P_P2	P_P3	–	–
number of items	3	3	3	–	–
L (18 itens)					
name of parcels	L_P1	L_P2	L_P3	L_P4	L_P5
number of items	3	3	4	4	4

Note. N = Neuroticism. E = Extraversion. P = Psychoticism. L = Lie/Social Desirability.

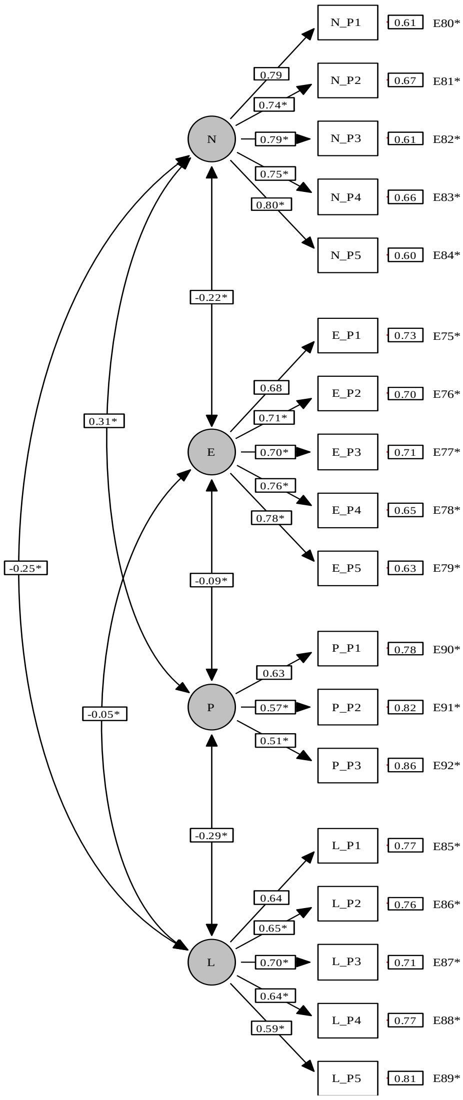

Figure 1

Path diagram of the EPQ-R confirmatory factor analysis model.

Table 5

Measurement invariance analysis.

	CFI	SRMR	RMSEA	χ^2	df	Δdf	$\Delta\chi^2$	ΔCFI
Configural	.961	.043	.031	665.363	258			
Metric	.960	.044	.030	684.113	272	14	18.750, p=.175	.001
Scalar	.935	.046	.037	969.560	290	32	304.196, p<.001	.026
Scalar (partial)	.960	.044	.031	693.607	.277	19	28.243, p=.079	.001

Note. CFI = Comparative Fit Index. SRMR = Standardized Root Mean Square Residual. RMSEA= Root Mean Square Error of Approximation. χ^2 = chi-square. df = degrees of freedom. $\Delta\chi^2$, Δdf and ΔCFI were the difference between each alternative and the configural model.