Personality and Individual Differences 149 (2019) 108–117

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Development of a new abbreviated form of the Eysenck Personality Questionnaire-Revised with multidimensional item response theory

Daiana Colledani*, Pasquale Anselmi, Egidio Robusto

Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Italy

ARTICLE INFO

Keywords:
EPQ-R
PEN-L traits
Multidimensional IRT
Differential item functioning
Latent profile analysis

ABSTRACT

The present work consists of two studies. In the first study, a new abbreviated form of the EPQ-R (six items per scale) was developed from the 100 items of the full-length version of the questionnaire. Methods and procedures developed within the framework of multidimensional IRT were used for this purpose. In the second study, the abbreviated questionnaire was validated on a new data sample. In addition, latent profile analysis was used to identify groups of individuals characterized by similar patterns of the four PEN-L traits. These patterns were also compared with respect to indicators of psychosocial functioning. Results indicated that the new abbreviated form of the questionnaire outperforms the old abbreviated form with respect to reliability and approximation of measures obtained with the full-length test. Moreover, the four-factor structure of the instrument and its convergent validity have been confirmed. Three PEN-L patterns have been identified that differ for anxiety and depression, satisfaction for social relations, frequency of substance use and sexual risk behaviors.

1. Introduction

Eysenck's questionnaires are among the most used instruments for the assessment of personality (Eysenck & Eysenck, 1991). Over the years, several contributions have been offered for the refinement of these questionnaires and for the development of brief versions for both adult and young people (e.g., Francis, 1996; Francis, Brown, & Philipchalk, 1992; Francis & Pearson, 1988). The short forms of the Eysenck Personality Questionnaire-Revised (EPQ-R; Eysenck & Eysenck, 1991; Eysenck, Eysenck, & Barrett, 1985) assess the four PEN-L (Psychoticism, Extraversion, Neuroticism, Lie) traits through 48 items (12 per scale; Corulla, 1990; Eysenck, Eysenck, & Barrett, 1985; Francis & Pearson, 1988). Although the short forms of the EPQ-R were explicitly developed "for use when time is very limited" (Eysenck, Pearson, Easting, & Allsopp, 1985; p. 24), it has been argued that, in some cases, these forms could still be too long, thus leading researchers to exclude the assessment of some personality traits (Francis et al., 1992). Consequently, abbreviated forms of the EPQ-R have been developed that include 24 items only (6 per scale; Francis, 1996; Francis et al., 1992).

In general, research provided evidence about the cross-cultural validity of these instruments and their acceptable psychometric properties. However, some criticisms have also been raised. For instance, concerning the abbreviated form of the EPQ-R (Francis et al., 1992), different studies highlighted not fully satisfactory reliability, mainly for P and L scales (Forrest, Lewis, & Shevlin, 2000; Shevlin, Bailey, & Adamson, 2002). In addition, other studies showed that several items of P, N, and L scales might exhibit differential item functioning (DIF) across gender (e.g., Escorial & Navas, 2007; Forrest et al., 2000; Karanci, Dirik, & Yorulmaz, 2007), which makes the comparison between groups questionable.

The present work consists of two studies. The first study aims at developing a new version of the abbreviated form (i.e., that consisting of six items for each scale) of the EPQ-R for adults, with improved psychometric properties. To this purpose, the 24 items with the best psychometric properties are selected from the full-length version of the questionnaire using statistics and procedures developed within the framework of item response theory (IRT; Bock, 1997; Thissen & Steinberg, 2009).

IRT provides useful information concerning the psychometric properties of the items (Bortolotti, Tezza, de Andrade, Bornia, & de Sousa Júnior, 2013; Petrillo, Cano, McLeod, & Coon, 2015; Spence, Owens, & Goodyer, 2012) and has been found to be effective for the development and validation of measurement scales (see, e.g., Anselmi, Vianello, Voci, & Robusto, 2013; Anselmi, Vidotto, Bettinardi, & Bertolotti, 2015; Balsamo, Giampaglia, & Saggino, 2014; Colledani, 2018, Da Dalt et al., 2013, 2015, 2017; Sotgiu, Anselmi, Meneghini, in press; Vidotto, Anselmi, Filipponi, Tommasi, & Saggino, 2018; Zanon,

* Corresponding author at: Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Via Venezia 14, 35131, Padova, Italy.
E-mail address: daiana.colledani@unipd.it (D. Colledani).

https://doi.org/10.1016/j.paid.2019.05.044
Received 30 January 2019; Received in revised form 21 May 2019; Accepted 26 May 2019
Available online 05 June 2019
0191-8869/

Hutz, Yoo, & Hambleton, 2016). IRT, in particular, allows for identifying the items that, while covering the entire latent trait continua, are best at discriminating different levels of the traits. The selection of these items leads to short scales that produce scores very similar to those obtained with the full-length versions of the instruments and with the same external validity (i.e., the same correlations with other constructs; Reise & Henson, 2000; Spence et al., 2012). Moreover, IRT allows for detecting items that are unclear, ambiguous, or which exhibit DIF. Recently, Colledani, Robusto, and Anselmi (2018) and Colledani, Anselmi, and Robusto (2018) used IRT models for developing a new abbreviated form of the Junior EPQ-R and a new short form of the EPQ-R, respectively. The new forms outperformed the old ones on several aspects.

Multidimensional IRT (MIRT) models (see, e.g., Haberman, von Davier, & Lee, 2008; Reckase, 2009) are used in this work. These models offer the same advantages of unidimensional IRT models, plus others that are particularly useful for the analysis of multidimensional questionnaires like the EPQ-R. In particular, MIRT models allow for better understanding the traits measured by an instrument and how they are related to each other. Moreover, these models allow for identifying the contribution of individual items to the measurement of each trait (Ackerman, 1994). As a further aspect, MIRT models could provide a more precise estimation of scale reliability (Cheng, Wang, & Ho, 2009), item parameters (Finch, 2010), and person parameters (van den Berg, Paap, Derks, & Outcome of Psychosis (GROUP) investigators, 2013). Recently, Colledani, Anselmi, and Robusto (2019) used a MIRT model for developing an abbreviated form of the Eysenck's Impulsiveness-Venturesomeness-Empathy Questionnaire (Eysenck, Pearson, et al., 1985). Reliability, validity, and trait scores of the abbreviated form largely resembled those of the full-length version of the questionnaire.

The second study aims at validating the new abbreviated form of the EPQ-R developed in Study 1 on a new data sample. In addition, a Latent Profile Analysis (LPA) was conducted. LPA is a person-centered analysis that aims at identifying groups of individuals characterized by similar patterns of scores on a set of variables (Collins & Lanza, 2010). In personality psychology, LPA may be used to define higher-order personality typologies which allow for describing personality better than individual scale scores (Ferguson & Hull, 2018; Kinnunen et al., 2012; Merz & Roesch, 2011; Parr, Lanza, & Bernthal, 2016; Zhang, Bray, Zhang, & Lanza, 2015). In the present study, LPA has been used to identify groups of individuals with similar patterns of the four PEN-L traits. In addition, the identified patterns were compared with respect to indicators of psychosocial functioning, such as anxiety and depression, satisfaction for social relations, frequency of substance use and sexual risk behaviors.

2. Study 1

2.1. Method

2.1.1. Participants, material, and procedure

A total of 570 native Italian speakers (females = 331; mean age = 28.73, *SD* = 11.87, from 18 to 84 years) were recruited through convenience sampling. All participants completed, anonymously and voluntarily, the Italian version (Dazzi, Pedrabissi, & Santinello, 2004) of the EPQ-R (Eysenck, Eysenck, & Barrett, 1985). An instruction letter given before completion asked them to be honest and quick in their answers. All the standards for research with human subjects were respected, and only gender and age were asked.

The EPQ-R consists of 100 dichotomous items (yes/no), 32 for P scale (e.g., “Should people always respect the law?”, “Do you enjoy hurting people you love?”), 23 for E scale (e.g., “Do you enjoy meeting new people?”, “Can you get a party going?”), 24 for N scale (e.g., “Would you call yourself a nervous person”, “Are you often troubled about feelings of guilt?”), and 21 for L scale (e.g., “Are all your habits good and desirable ones?”, “Have you ever cheated at a game?”). The Italian version of the questionnaire showed adequate reliability for all scales (Cronbach's α = 0.67, 0.78, 0.85, and 0.75 for P, E, N, and L scales, respectively; Dazzi et al., 2004), and the four-factor structure was confirmed (Dazzi, 2011).

2.1.2. Analysis strategy

The two-parameter logistic (2PL) MIRT model was estimated on the responses to the 100 items of the full-length version of the EPQ-R through the R-package “mirt” (Chalmers et al., 2018). In the 2PL MIRT model, the probability that an individual endorses a certain item (i.e., provides a “yes” response to that item) is a function of: a) the parameter θ of the individual, which represents the level of the individual on the latent trait; b) the parameter ε of the item, which defines the “endorsability” of the item (i.e., the ease of providing a “yes” answer to it); and c) the parameter δ of the item, which expresses the capability of the item in discriminating individuals with different trait levels. For each individual and each item, there are as many parameters θ and δ as the latent traits that are measured by the questionnaire. In this work, the exploratory 2PL MIRT model was estimated considering four factors, one for each of the four scales of the EPQ-R.

Selection of the 24 items with the best psychometric properties was accomplished by taking into account discrimination (δ) and easiness (ε) parameters of each item, gender DIF, and item misfit. Items with low discrimination parameter in the intended trait or with large discrimination parameters in more than one trait (complex structured items) were not selected for inclusion in the abbreviated questionnaire. Research (see, e.g., Luecht & Miller, 1992; Yao & Boughton, 2009), in fact, has shown that these items are generally non-informative about individual's trait level and supports the use of simple structured items (i.e., items with large discrimination parameter in one trait only). For instance, Mulder and van der Linden (2009) found that trait estimates are more accurate when simple structured items are used. Sinharay, Puhan, and Haberman (2011) found that scores of different traits must be sufficiently distinct from each other to have adequate psychometric quality. This condition is more easily satisfied if there are no items measuring more than one trait. The signed chi-squared test (S-χ^2; Orlando & Thissen, 2000) was used to identify misfitting items, which were not selected for inclusion in the abbreviated questionnaire. A multiple-group confirmatory 2PL MIRT model was used to detect gender DIF. Invariance of easiness (uniform bias) and discrimination (non-uniform bias) parameters was explored through the Wald test. Items exhibiting uniform or non-uniform biases of medium ($0.3 \leq \Phi < 0.5$) to large ($\Phi \geq 0.5$) size (Cohen, 1988) were not selected for inclusion in the abbreviated questionnaire. From the remaining items, those that allowed for better covering the entire continuum of their own latent trait and with the largest discrimination level were chosen. Six items were selected for each scale.

Reliability of the new abbreviated scales was evaluated through Cronbach's α. A bias index was computed as the average difference (in absolute terms) between the latent trait estimates obtained on the full-length scales and those obtained on the abbreviated scales. Low biases suggest that the latent trait estimates obtained with the abbreviated scales approximate those resulting from the full-length versions. In addition, Pearson's correlation coefficients between the (raw) scores obtained on the abbreviated scales and those obtained on the full-length versions were computed, with the correction for common items suggested by Levy (1967).

2.2. Results

Parameter estimates of the items of the P, E, N, and L scales are reported in Tables 1, 2, 3, and 4. Descriptive statistics, reliability coefficients, bias indices, and corrected correlations between the scores of the abbreviated and full-length scales are reported in Table 5. In the tables, “Old” denotes the existing abbreviated form of the EPQ-R

Table 1
DIF statistics, item fit, and MIRT parameters for the 32 items of Psychoticism scale. The items are ordered by increasing easiness (parameter ε). The items included in the old abbreviated scales and in the new abbreviated scales are marked by "✓".

Item	Old	New	δ_P	δ_E	δ_N	δ_L	E	S-χ^2	*df*	DIF-δ	DIF-ε
Epq-r30			−1.88	0.75	0.52	0.11	−5.26	11.77	6.00	0.04	0.01
Epq-r12			−2.08	0.57	−0.62	−0.07	−3.65	30.80	20.00	0.03	0.11
Epq-r37		✓	−1.44	−0.20	0.25	−0.84	−3.64	25.18	18.00	0.07	0.13
Epq-r96		✓	−1.58	0.41	−0.61	−0.72	−3.32	23.18	20.00	0.07	0.03
Epq-r56			−0.59	0.21	0.66	0.16	−3.09	19.94	16.00	0.03	0.05
Epq-r54			−1.95	1.41	0.34	0.06	−3.02	56.56***	25.00	0.05	0.06
Epq-r25	✓		−0.67	−0.20	0.20	−0.94	−2.68	24.39	20.00	0.15	0.14
Epq-r73			−0.47	0.38	0.92	0.16	−2.64	25.57	20.00	0.00	0.06
Epq-r68			−1.11	0.19	−0.06	0.11	−2.55	22.42	22.00	0.03	0.05
Epq-r79			−0.86	0.23	−0.03	0.02	−2.49	19.12	21.00	0.04	0.04
Epq-r34			−1.00	−0.29	0.99	0.47	−2.37	26.06	20.00	0.03	0.05
Epq-r41		✓	−1.38	0.32	−0.35	−0.82	−2.36	32.74	23.00	0.01	0.07
Epq-r59	✓		−1.45	−0.20	−0.54	0.15	−2.31	32.40	22.00	0.05	0.13
Epq-r48	✓		−0.97	0.15	0.11	0.17	−2.00	17.14	24.00	0.08	0.15
Epq-r7	✓		−1.25	0.02	−0.71	−0.14	−1.73	42.1*	25.00	0.04	0.23
Epq-r21			−0.88	0.06	−0.36	−0.72	−1.71	23.96	26.00	0.03	0.11
Epq-r99			−0.51	0.11	−0.01	−0.03	−1.52	27.21	25.00	0.08	0.10
Epq-r91			−0.48	0.04	0.72	−0.20	−1.49	74.14***	24.00	0.03	0.03
Epq-r81		✓	−1.29	−0.40	0.43	−0.37	−1.45	26.89	21.00	0.04	0.00
Epq-r95			−0.46	−0.05	1.07	0.58	−1.21	36.11	24.00	0.01	0.04
Epq-r2			−1.19	−0.47	0.30	−0.29	−1.18	28.58	23.00	0.02	0.02
Epq-r14			−0.51	−0.01	−0.47	−0.17	−0.88	37.45	28.00	0.02	0.22
Epq-r5			−0.85	−0.24	−0.26	0.33	−0.65	46.5**	27.00	0.02	0.12
Epq-r64			−0.21	0.00	0.08	−0.52	−0.62	52.56**	28.00	0.01	0.10
Epq-r75			−0.29	−0.15	−0.07	−0.08	−0.53	26.95	27.00	0.03	0.15
Epq-r42		✓	−0.81	−0.12	0.26	−0.44	−0.38	32.72	25.00	0.05	0.07
Epq-r18			−0.70	−0.14	−0.17	−0.88	−0.35	26.01	28.00	0.01	0.10
Epq-r50			−0.15	−0.11	0.37	0.44	0.04	31.65	27.00	0.09	0.04
Epq-r29	✓	✓	−1.08	−0.16	0.21	−0.49	0.33	30.13	24.00	0.05	0.09
Epq-r88	✓		−0.37	−0.06	0.05	−0.58	0.34	56.15***	27.00	0.03	0.09
Epq-r9			−0.15	0.14	−0.09	−0.37	0.54	36.32	28.00	0.12	0.04
Epq-r85			−0.67	0.23	0.60	−0.06	0.58	22.19	25.00	0.11	0.02

Note. δ_P, δ_E, δ_N, δ_L = discrimination parameter for P, E, N, and L scales, respectively. S-χ^2 = item fit index; *df* = degrees of freedom of S-χ^2; DIF-δ = effect size of non-uniform bias; DIF-ε = effect size of uniform bias.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 2
DIF statistics, item fit, and MIRT parameters for the 23 items of Extraversion scale. The items are ordered by increasing easiness (parameter ε). The items included in the old abbreviated scales and in the new abbreviated scales are marked by "✓".

Item	Old	New	δ_P	δ_E	δ_N	δ_L	E	S-χ^2	*df*	DIF-δ	DIF-ε
Epq-r90		✓	−0.47	−1.02	0.15	−0.32	−2.04	22.32	22.00	0.00	0.02
Epq-r61			−0.63	−0.96	1.01	−0.43	−1.01	34.46*	21.00	0.02	0.06
Epq-r47	✓		−0.02	−0.10	0.06	−0.29	−0.38	35.76	28.00	0.10	0.20
Epq-r40			−0.40	−0.66	−0.67	0.02	−0.26	23.09	28.00	0.06	0.14
Epq-r1			−0.07	−0.75	0.04	−0.03	−0.07	29.17	26.00	0.02	0.20
Epq-r51	✓	✓	−0.32	−2.18	0.22	0.05	0.08	34.27*	21.00	0.02	0.15
Epq-r67			−0.68	−0.71	0.14	0.38	0.15	41.88*	24.00	0.13	0.15
Epq-r69			−0.69	−0.83	0.97	0.21	0.20	30.53	20.00	0.09	0.02
Epq-r63			−0.35	−0.46	0.04	0.05	0.22	34.65	26.00	0.01	0.11
Epq-r72			0.06	−0.40	0.34	−0.09	0.24	29.40	26.00	0.05	0.09
Epq-r45		✓	−0.11	−1.67	0.15	0.33	0.28	32.45	24.00	0.07	0.03
Epq-r55			−0.11	−0.96	−0.10	−0.16	0.59	27.42	25.00	0.02	0.21
Epq-r24	✓		0.49	−1.58	−0.62	0.01	0.83	36.43	28.00	0.40	0.02
Epq-r6	✓	✓	0.07	−1.14	−0.01	0.27	1.07	28.51	25.00	0.02	0.01
Epq-r94	✓		−0.12	−1.17	−0.11	0.65	1.14	31.92	25.00	0.05	0.14
Epq-r36			0.03	−1.46	−0.50	−0.32	1.23	18.38	25.00	0.07	0.23
Epq-r78			0.21	−1.06	0.12	0.02	1.26	42.68*	24.00	0.04	0.00
Epq-r28			0.16	−1.45	0.27	−0.30	1.69	21.88	23.00	0.04	0.08
Epq-r33			0.78	−1.40	−0.22	−0.58	1.76	60.88***	26.00	0.01	0.03
Epq-r16		✓	0.42	−1.74	−0.48	−0.53	2.03	31.55	25.00	0.07	0.06
Epq-r11	✓		0.73	−1.74	−0.37	0.28	3.69	22.01	18.00	0.03	0.01
Epq-r20			1.52	−2.20	−0.17	0.26	3.97	32.29	22.00	0.03	0.01
Epq-r58		✓	1.55	−2.60	−0.29	−0.08	4.13	27.23	23.00	0.03	0.03

Note. δ_P, δ_E, δ_N, δ_L = discrimination parameter for P, E, N, and L scales, respectively. S-χ^2 = item fit index; *df* = degrees of freedom of S-χ^2; DIF-δ = effect size of non-uniform bias; DIF-ε = effect size of uniform bias.

* $p < .05$.

*** $p < .001$.

Table 3
DIF statistics, item fit, and MIRT parameters for the 24 items of Neuroticism scale. The items are ordered by increasing easiness (parameter ε). The items included in the old abbreviated scales and in the new abbreviated scales are marked by "✓".

Item	Old	New	δ_P	δ_E	δ_N	δ_L	E	S-χ^2	*df*	DIF-δ	DIF-ε
Epq-r38	✓		−0.837	−0.053	1.086	−0.355	−3.009	34.80*	19	0.049	0.090
Epq-r70		✓	−0.43	0.519	1.409	−0.21	−2.038	33.903	24	0.036	0.022
Epq-r60			−0.18	−0.083	0.599	0.247	−1.668	23.304	23	0.013	0.048
Epq-r26	✓	✓	−0.076	0.369	1.674	−0.093	−1.25	39.544*	25	0.057	0.084
Epq-r46		✓	0.138	0.363	2.534	0.227	−1.147	45.01**	23	0.141	0.053
Epq-r35	✓		−0.001	0.055	2.702	0.124	−0.863	37.08*	22	0.000	0.092
Epq-r76			−0.337	0.389	0.68	−0.204	−0.825	23.153	27	0.028	0.088
Epq-r84	✓		0.041	0.047	0.862	0.163	−0.563	39.304	27	0.008	0.080
Epq-r83	✓		−0.048	−0.086	1.368	0.353	−0.477	45.54**	23	0.033	0.073
Epq-r17		✓	0.068	0.145	1.944	−0.359	−0.461	47.8**	24	0.026	0.081
Epq-r8			0.255	0.259	0.984	−0.457	−0.291	37.984	28	0.022	0.221
Epq-r31		✓	0.65	0.154	1.335	−0.427	−0.291	42.02*	28	0.099	0.210
Epq-r100			0.009	−0.193	0.84	−0.026	−0.18	30.947	25	0.019	0.056
Epq-r65			0.302	0.12	0.847	−0.428	0.075	33.867	27	0.037	0.142
Epq-r3	✓	✓	0.303	−0.321	2.005	−0.399	0.191	31.114	22	0.086	0.186
Epq-r43			0.685	−0.124	1.024	0.187	0.312	47.91**	27	0.021	0.130
Epq-r74			0.261	−0.103	0.594	−0.116	0.348	29.819	27	0.042	0.102
Epq-r97			0.629	−0.172	0.572	−0.4	0.699	42.36*	28	0.090	0.137
Epq-r87			0.918	−0.041	0.423	−0.242	0.79	32.785	29	0.125	0.091
Epq-r92			0.212	−0.18	0.402	−0.546	0.826	36.304	27	0.050	0.000
Epq-r22			0.561	0.052	1.133	0.153	1.079	37.175	26	0.060	0.198
Epq-r80			0.867	0.547	0.87	−0.272	1.243	28.796	27	0.057	0.142
Epq-r13			0.844	−0.016	1.444	−0.009	1.313	41.214*	25	0.019	0.166
Epq-r52			0.694	−0.323	−0.035	0.013	1.585	24.69	27	0.057	0.015

Note. δ_P, δ_E, δ_N, δ_L = discrimination parameter for P, E, N, and L scales, respectively. S-χ^2 = item fit index; *df* = degrees of freedom of S-χ^2; DIF-δ = effect size of non-uniform bias; DIF-ε = effect size of uniform bias.
* $p < .05$.
** $p < .01$.

(Francis et al., 1992), "New" denotes the abbreviated form developed in the present work.

2.2.1. Psychoticism scale

Out of the 32 items of this scale, 15 exhibited complex structure and/or low discrimination in the intended trait (Items 9, 14, 18, 21, 25, 34, 50, 56, 54, 64, 73, 85, 88, 91, and 95) and six items showed moderate misfit (Items 5, 7, 54, 64, 88, and 91). No item exhibited uniform or non-uniform gender DIF. Three of these poor-functioning items (Items 7, 25, and 88) were included in the old abbreviated scale.

Among the 15 remaining items, six were selected taking into account easiness and discrimination parameters. The new scale consists of Items 29, 37, 41, 42, 81, and 96. Only Item 29 was present also in the old abbreviated version. The new abbreviated scale outperformed the

Table 4
DIF statistics, item fit, and MIRT parameters for the 21 items of Lie scale. The items are ordered by increasing easiness (parameter ε). The items included in the old abbreviated scales and in the new abbreviated scales are marked by "✓".

Item	Old	New	δ_P	δ_E	δ_N	δ_L	E	S-χ^2	*df*	DIF-δ	DIF-ε
Epq-r53			−0.715	0.516	−0.438	1.152	−1.861	60.38***	25	0.110	0.065
Epq-r93			0.201	−0.091	−0.375	0.413	−1.597	54.91***	25	0.144	0.050
Epq-r23		✓	0.099	−0.202	−0.075	0.953	−1.067	30.846	26	0.035	0.045
Epq-r10	✓	✓	−0.347	0.069	0.366	1.318	−1.033	45.76**	26	0.031	0.133
Epq-r82			0.159	0.442	−0.266	0.32	−0.743	44.31*	27	0.070	0.036
Epq-r89			−0.168	0.296	−0.259	0.57	−0.653	45.85*	29	0.036	0.043
Epq-r39			0.232	0.082	0.043	0.482	−0.619	19.329	27	0.002	0.014
Epq-r44			−0.013	0.258	−0.082	0.844	−0.563	41.435	29	0.051	0.061
Epq-r86	✓	✓	−0.28	−0.134	0.09	1.004	−0.556	24.505	27	0.108	0.076
Epq-r27	✓	✓	0.268	−0.097	0.065	1.105	−0.46	41.83*	28	0.033	0.058
Epq-r49			0.005	0.442	0.122	0.448	−0.407	44.23	29	0.138	0.045
Epq-r66	✓	✓	−0.015	0.409	0.171	1.385	−0.354	37.909	29	0.049	0.005
Epq-r77			0.547	0.301	0.077	0.185	−0.113	33.85	29	0.154	0.007
Epq-r32			0.295	−0.072	−0.02	0.731	0.149	44.58*	29	0.138	0.040
Epq-r15			−0.049	−0.226	0.081	1.215	0.217	34.01	28	0.032	0.008
Epq-r71	✓	✓	0.457	−0.237	−0.036	1.299	0.353	40.778	29	0.008	0.053
Epq-r62			0.215	−0.352	0.131	0.564	0.457	29.185	28	0.133	0.156
Epq-r19	✓		0.307	−0.142	−0.204	0.778	1.08	37.909	28	0.027	0.017
Epq-r4			0.469	−0.13	−0.104	0.791	1.174	27.838	28	0.091	0.038
Epq-r57			1.278	−0.087	−0.279	0.074	1.564	41.55*	26	0.082	0.032
Epq-r98			0.289	−0.223	−0.684	0.668	2.05	32.662	28	0.025	0.093

Note. δ_P, δ_E, δ_N, δ_L = discrimination parameter for P, E, N, and L scales, respectively. S-χ^2 = item fit index; *df* = degrees of freedom of S-χ^2; DIF-δ = effect size of non-uniform bias; DIF-ε = effect size of uniform bias.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5
Descriptive statistics, reliability coefficients, bias indices, and correlations (between abbreviated and full-length scales) for the old abbreviated scales and the new abbreviated scales.

	Psychoticism		Extraversion		Neuroticism		Lie	
	Old	New	Old	New	Old	New	Old	New
Mean	1.78	1.61	3.86	3.6	2.17	2.33	2.77	2.36
SD	1.28	1.33	1.55	1.55	1.66	1.85	1.66	1.67
Bias	0.47	0.37	0.46	0.32	0.37	0.32	0.36	0.33
Cronbach's α	0.47	0.54	0.59	0.65	0.65	0.72	0.60	0.61
Cronbach's α 95% CI	0.39, 0.52	0.48, 0.59	0.54, 0.64	0.61, 0.69	0.61, 0.70	0.69, 0.76	0.54, 0.65	0.56, 0.66
Correlation corrected for common items	0.56	0.67	0.61	0.71	0.68	0.75	0.65	0.65

old one in all the considered indices. Reliability and Levy's corrected correlation increased, whereas bias between the scores obtained on the abbreviated and full-length scales decreased (see Table 5).

2.2.2. Extraversion scale

In this scale, eight items showed complex structure or low discrimination in the intended trait (Items 61, 47, 40, 67, 69, 63, 72 and 20). In addition, three items exhibited moderate misfit (Items 67, 33, and 78) and one item showed a moderate non-uniform gender DIF (Item 24). Two of these poor-functioning items (Items 24 and 47) were included in the old abbreviated scale.

The six items for the abbreviated scale were selected among the remaining 12 items on the basis of the ε and δ parameters. The resulting scale (Items 90, 51, 45, 6, 16, and 58) includes only two items which were present also in the old abbreviated version (Items 51 and 6). Compared with the old abbreviated scale, the new one showed larger reliability and Levy's corrected correlation, and lower bias between the scores obtained on the abbreviated and full-length scales (see Table 5).

2.2.3. Neuroticism scale

In this scale, seven items showed a complex structure (Items 38, 43, and 80) or low discrimination in the intended trait (Items 97, 87, 92, and 52), whereas no item exhibited misfit or gender DIF. One of these poor-functioning items (Item 38) was included in the old abbreviated scale.

Easiness and discrimination parameters of the remaining 17 items were inspected to select the six items required to compose the abbreviated scale. The resulting scale (Items 3, 17, 26, 31, 46, and 70) includes only two items (Items 3 and 26) in common with the old abbreviated scale. Compared with the old abbreviated scale, the new one showed larger reliability and Levy's corrected correlation, and lower bias between the scores obtained on the abbreviated and full-length scales (see Table 5).

2.2.4. Lie scale

Out of the 21 items of this scale, two exhibited complex structure (Items 98 and 53) and five showed low discrimination in the intended trait (Items 93, 82, 49, 77, and 57). One item exhibited moderate to strong misfit (Item 53) and no item showed gender DIF.

Among the 14 remaining items, six were selected considering their location on the latent trait and their discrimination parameters. The new version of the scale consists of Items 23, 10, 86, 27, 66, and 71. Only Item 23 is not present in the old abbreviated version. The new abbreviated scale, obtained replacing only one item, showed slightly larger reliability and slightly lower bias (see Table 5). Levy's corrected correlation remained unchanged.

It is worth noting that the short form (12 items per scale) of the EPQ-R that has been developed by Colledani, Anselmi, and Robusto (2018; Items 18, 21, 25, 29, 30, 37, 41, 42, 64, 81, 88, 96 for P scale; Items 6, 16, 20, 24, 28, 45, 51, 58, 61, 67, 78, 90 for E scale; Items 3, 17, 22, 26, 31, 38, 46, 65, 70, 80, 83, 84 for N scale; Items 4, 10, 19, 23, 27, 32, 44, 53, 66, 71, 86, 93 for L scale) include the abbreviated form that has been developed in the present study. Correlations between the short and abbreviated forms are strong (r = 0.88, 0.90, 0.92, and.89 for P, E, N, and L, respectively).

2.3. Discussion

This study aimed to develop a new version of the abbreviated form of the EPQ-R, with improved psychometric properties. To this purpose, several item-level analyses were performed and some items of the old abbreviated scales were replaced with items better located on the latent traits, with higher discrimination, and without gender DIF, misfit, or complex structure. The new abbreviated scales differed from the old ones for four (E and N scales) to five (P scale) items, whereas only one item was changed on L scale.

Overall, the new abbreviated scales outperformed the old ones in all the considered indices. Compared with the old abbreviated scales, the new ones showed larger reliability. In addition, compared with the old abbreviated scales, the new ones showed stronger Levy's corrected correlation between the (raw) scores computed on the abbreviated and full-length scales (only for L scale, Levy's corrected correlation remained unchanged), as well as lower bias between the latent trait estimates obtained on the abbreviated and full-length scales.

3. Study 2

This study aims at validating the abbreviated form of the EPQ-R developed in Study 1 on a new data sample. Construct validity is tested by analyzing the factor structure of the instrument and by exploring correlations between the abbreviated scales and the dimensions of the five-factor model (FFM) of personality.

Furthermore, LPA is used to identify groups of individuals showing similar patterns of the four PEN-L traits. The identified patterns are compared with respect to a series of indicators of psychosocial functioning, such as anxiety, depression, use of substances, sexual risk behaviors, and satisfaction for social relations.

3.1. Method

3.1.1. Participants, materials, and procedure

A total of 303 native Italian speakers (females = 71.2%; mean age = 27.41, *SD* = 9.09, from 18 to 65 years) were recruited through convenience sampling. The participation to the study was anonymous and voluntary, and all standards for research with human subjects were respected.

All participants were presented with the abbreviated form of the EPQ-R developed in Study 1 (in this sample, Cronbach's αs = 0.53, 0.69, 0.70, and 0.61 for P, E, N, and L scales, respectively; Cronbach's αs 95% CIs = 0.46–0.61, 0.63–0.74, 0.66–0.76, and 0.55–0.68 for P, E, N, and L scales, respectively). In addition, a sub-sample of participants (N = 206, females = 76.0%; mean age = 27.49, *SD* = 9.95, from 18 to 65 years) completed other measures aimed to evaluate the five traits of the FFM of personality, anxiety and depression, satisfaction for social

relations, frequency of use of substances and sexual risk behaviors.

The Italian version (Chiorri, Marsh, Ubbiali, & Donati, 2016; Ubbiali, Chiorri, & Hampton, 2013) of the Big Five Inventory (BFI; John, Naumann, & Soto, 2008) was used to evaluate the five traits of the FFM. The questionnaire consists of 44 items answered on a five-point Likert scale (from 1 "Strongly disagree" to 5 "Strongly agree"). Examples of items are: "I see myself as someone who is talkative" for extraversion; "I see myself as someone who likes to cooperate with others" for agreeableness; "I see myself as someone who is a reliable worker" for conscientiousness; "I see myself as someone who can be tense" for emotional stability; "I see myself as someone who has an active imagination" for openness. Construct validity, factor structure, and gender invariance of the BFI were supported (Ubbiali et al., 2013; Chiorri et al., 2016; Cronbach's αs from 0.75 to 0.86; in this sample, Cronbach's αs from 0.72 to 0.84).

The Italian version of the trait scale of the State-Trait Anxiety Inventory (STAI-Y; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983; Pedrabissi & Santinello, 1989) was used to evaluate anxiety. The scale assesses feelings of general anxiety and the relatively stable predisposition to view stressful situations as threatening. The scale includes 20 self-report items (e.g., "I am regretful") answered on a four-point Likert scale (from 1 "Not at all" to 4 "Very much"). The Italian version of the questionnaire showed adequate validity and reliability (Pedrabissi & Santinello, 1989; Cronbach's αs from 0.85 to 0.90; in this sample, Cronbach's $\alpha = 0.92$).

To evaluate depression, the Italian version of the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001; Spitzer, Kroenke, Williams, & Patient Health Questionnaire Primary Care Study Group, 1999) was used. The questionnaire is a self-report instrument including nine items (e.g., "Feeling tired or having little energy") intended to evaluate the nine DSM-IV (American Psychiatric Association, 2000) criteria for depression. Respondents are asked to evaluate the presence of depressive symptoms over the last 2 weeks on a four-point Likert scale (from 0 "Not at all" to 3 "Nearly every day"). The instrument showed adequate reliability, and good sensitivity and specificity (Kroenke et al., 2001; Cronbach's αs from 0.86 to 0.89; in the current sample, Cronbach's $\alpha = 0.83$).

Four items were administered to evaluate satisfaction for social relations ("I feel satisfied about my social relationships", "I often wish to have more social contacts and closeness with people", "My social life is rich and fulfilling", "In general, I feel that my loved ones are close to me"). The items were scored on a five-point Likert scale (from 1 "Strongly disagree" to 5 "Strongly agree"). In the current sample, Cronbach's $\alpha = 0.77$.

Finally, six self-report items were used to assess the frequency of substance use and sexual risk behaviors in the last six months. The six items were answered on a five-point Likert scale (from 1 "Never" to 5 "Very often"). Three of these items evaluated the frequency of drug consumption. The responses to these three items were averaged to obtain a single score (Cronbach's $\alpha = 0.93$). The remaining three items evaluated the frequency of drinking alcohol, smoking tobacco, and having unprotected sexual intercourse.

3.1.2. Analysis strategy

Construct validity of the new abbreviated form of the EPQ-R was evaluated through confirmatory factor analysis (CFA) and by computing convergent validity coefficients. CFA was conducted using the 24 items of the questionnaire as indicators (six for each factor). WLSMV (weighted least squares mean and variance-adjusted; Muthén & Muthén, 2012) was used as estimator. This method, in fact, is recommended for models involving categorical observed data (Brown, 2006; Flora & Curran, 2004) such as those of the EPQ-R. The goodness of fit of the model was evaluated using χ^2 statistics, Comparative Fit Index (CFI; Bentler, 1990), Weighted Root Mean Square Residual (WRMR; Muthén & Muthén, 2012), and Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993). A solution fits the data well when χ^2 is non-significant ($p \geq .05$), CFI is above 0.95 (0.90 to 0.95 for acceptable fit; Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004), RMSEA is smaller than 0.06 (between 0.06 and 0.08 for moderate fit; Browne & Cudeck, 1993), and WRMR is close to 1.0 (Yu, 2002).

Convergent validity was evaluated considering correlations between the traits measured by the EPQ-R and the dimensions of the FFM, as measured by the BFI. According with literature, a positive correlation was expected between E scores and the extraversion measure of the BFI. Positive correlations were expected also between N scores and the neuroticism measure of the BFI, while negative correlations were expected between the P scale and agreeableness and conscientiousness measures of the BFI (e.g., Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003; Draycott & Kline, 1995; Heaven, Ciarrochi, Leeson, & Barkus, 2013; McCrae & Costa, 1985; Saggino, 2000; Scholte & De Bruyn, 2004).

LPA was estimated on the responses to the abbreviated form of the EPQ-R. The analysis was conducted on the average scores obtained by each participant on the four traits measured by the questionnaire. Models with 1 to 5 latent classes were estimated and compared. In order to identify the best fitting model, several statistics were taken into account: Bayesian Information Criterion (BIC; Schwarz, 1978), Sample-Adjusted BIC (SABIC; Sclove, 1987), Akaike Information Criterion (AIC; Akaike, 1973), and Entropy. Concerning AIC, BIC, and SABIC, the lower the value, the better the fit (Nylund, Asparouhov, & Muthén, 2007). Entropy defines how well a model classifies individuals into the derived patterns. Entropy ranges from 0 to 1. The closer the value to 1, the better the fit (Celeux & Soromenho, 1996). Additionally, consecutive models were compared using the Vuong–Lo–Mendell–Rubin likelihood ratio test (VLMR; Nylund et al., 2007) and the Lo–Mendell–Rubin likelihood ratio test (adjusted LMR; Lo, Mendell, & Rubin, 2001). These tests are used to compare a model with C latent classes against a model with $C - 1$ classes. Significant p-values indicate that the model with C classes fits the data better than the more parsimonious model (i.e., the model with $C - 1$ classes). On the contrary, non-significant p-values suggest to retain the more parsimonious model. In addition, following the suggestions in the literature (e.g., Collins & Lanza, 2010), the interpretability of the solution was also considered for choosing the model.

As a final step, the PEN-L patterns identified through LPA were compared with respect to a series of indicators of psychosocial functioning. Specifically, one-way ANOVAs (with Bonferroni-corrected post-hoc pairwise comparisons) were performed to explore the differences in trait anxiety and depression, satisfaction for social relations, frequency of substance use and unprotected sexual intercourse across individuals with different PEN-L patterns.

3.2. Results

3.2.1. Construct validity

Construct validity of the abbreviated form of the EPQ-R was supported: The results of the CFA confirmed the four-factor structure of the questionnaire ($\chi^2(246) = 302.100$, $p = .009$; CFI = 0.95; RMSEA = 0.03, [0.02, 0.04], $p = .99$; WRMR = 0.97) and showed that all items loaded on the intended factor (loadings ranging from 0.51 to 0.93; see Table 6).[1]

In addition, all convergent validity coefficients were coherent with expectations. In particular, as shown in Table 7, strong positive correlations were found between E scale and the measures of extraversion

[1] A CFA was conducted on the old short form of the EPQ-R (Francis et al., 1992), using data from Study 1. The four-factor structure of the old version of the questionnaire was not confirmed ($\chi^2(246) = 558.675$, $p < .001$; CFI = 0.82; RMSEA = 0.05, [0.04, 0.05], $p = .81$; WRMR = 1.41; loadings from 0.21 to 0.85).

Table 6
Confirmatory factor analysis.

Item	λ
Epq-r29	0.63
Epq-r96	0.67
Epq-r41	0.69
Epq-r42	0.73
Epq-r37	0.76
Epq-r81	0.58
Epq-r16	0.68
Epq-r51	0.66
Epq-r6	0.63
Epq-r58	0.93
Epq-r90	0.51
Epq-r45	0.64
Epq-r17	0.54
Epq-r46	0.66
Epq-r3	0.72
Epq-r26	0.74
Epq-r70	0.84
Epq-r31	0.63
Epq-r66	0.55
Epq-r27	0.68
Epq-r86	0.59
Epq-r23	0.58
Epq-r10	0.53
Epq-r71	0.68

	Extraversion	Neuroticism	Lie
Psychoticism	0.07	0.41***	−0.52***
Extraversion		−0.27***	−0.10
Neuroticism			−0.24**

Note. λ = factor loading. All factor loadings are significant with $p \leq .001$.
** $p \leq .01$.
*** $p \leq .001$.

Table 7
Correlations of the new Abbreviated PEN-L Scales with the BFI Traits.

	P	E	N	L
Extraversion	0.009	0.740***	−0.268***	0.104
Agreeableness	−0.248***	0.182**	−0.301***	0.207**
Conscientiousness	−0.353***	0.026	−0.290***	0.321***
Neuroticism	0.152*	−0.156*	0.680**	−0.062
Openness	0.136	0.197**	−0.075	0.059

Note.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

($r = 0.740$) and neuroticism ($r = 0.680$) of the BFI. Moreover, two moderate negative correlations were found between P scale and the measures of conscientiousness ($r = -0.353$) and agreeableness ($r = -0.248$) of the BFI.

3.2.2. Latent profile analysis

Table 8 provides fit indices of the LPA. The inspection of the values reported in the table suggests to consider the three-class model as the best fitting. Although the AIC and SABIC values decreased as the number of classes increased, Entropy and BIC suggested to choose the three-class model. Furthermore, the VLMR and LMR tests indicated that the three-class model should have been preferred to the model with two classes ($p < .001$ for VLRM and LMR) and to the model with four classes ($p = .374$ for VLMR; $p = .390$ for LMR).

Fig. 1 depicts the mean scores on the four PEN-L traits for the individuals belonging to the three classes. As shown in the figure, the individuals in the three classes did not differ with respect to E, whereas they differed for the other traits. The individuals belonging to Class 1

Table 8
Fit statistics for LPA models with 1 to 5 latent classes.

	Number of classes *C*				
	1	2	3	4	5
AIC	224.648	163.058	143.888	130.527	115.481
BIC	254.357	211.337	210.735	215.943	219.466
SABIC	228.986	170.108	153.649	142.999	130.664
Entropy		0.738	0.849	0.739	0.730
VLMR (*C* v. *C* − 1)		71.589	39.781	23.361	25.046
P		< 0.001	< 0.001	0.374	0.647
LMR (*C* v. *C* − 1)		69.168	38.436	22.571	24.199
P		< 0.001	< 0.001	0.390	0.653

Note. AIC: Akaike information criterion; BIC: Bayesian information criterion; SABIC: Sample-Adjusted BIC; VLMR: Vuong–Lo–Mendell–Rubin; LMR: Adjusted Lo–Mendell–Rubin.

(only 4.4% of the sample) were characterized by high values of P and N, and relatively low values of L. These individuals may be labelled as *problematic*. In contrast, the individuals belonging to Class 2 were characterized by low values of P and N, and high values of L. These individuals may be labelled as *well-adjusted*, and represented the majority (58.1%) of the sample. Finally, the individuals belonging to Class 3 (37.6% of the sample) showed relatively high values of N and P. These individuals may be labelled as *anxious/antisocial*.

The individuals of the three classes differed for trait anxiety (F (2, 202) = 6.04, $p < .01$; $\eta^2 = 0.06$), depression (F (2, 201) = 6.53, $p < .01$; $\eta^2 = 0.06$), satisfaction for social relations (F (2,196) = 7.70, $p \leq .001$; $\eta^2 = 0.07$), and frequency of drug consumption (F (2, 273) = 7.71, $p < .001$; $\eta^2 = 0.05$), alcohol (F (2, 175) = 3.27, $p < .01$; $\eta^2 = 0.05$) and tobacco use (F (2, 273) = 3.28, $p < .05$; $\eta^2 = 0.02$), and unprotected sexual intercourse (F (2, 201) = 6.53,

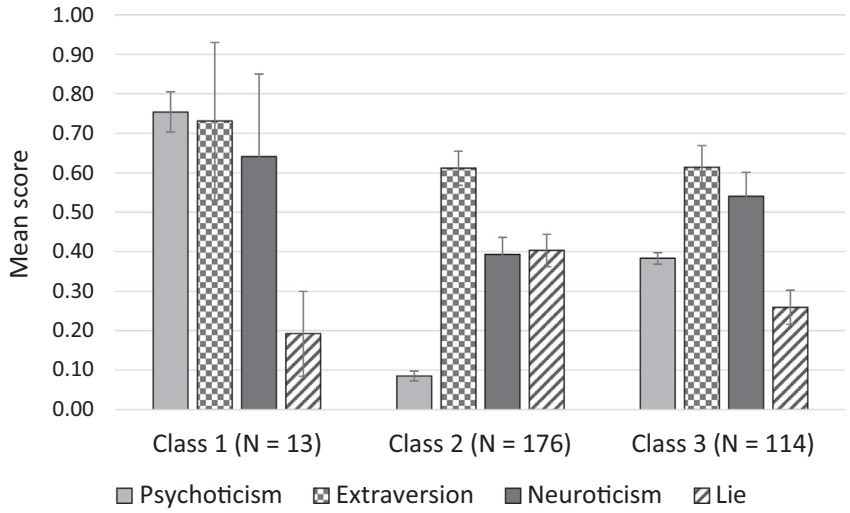

Fig. 1. Mean scores in the four PEN-L traits for the three classes resulting from the Latent Profile Analysis.

Table 9
Bonferroni-corrected post-hoc pairwise comparisons.

	Class	*N*	Mean	*SD*	Post-hoc pairwise comparisons
STAI	1	11	2.54	0.52	a*
	2	108	2.09	0.51	a*; b*
	3	84	2.29	0.56	b*
PHQ-9	1	11	1.12	0.58	a*
	2	107	0.71	0.43	a*; b**
	3	84	0.95	0.62	b**
Satisfaction for social relations	1	10	2.90	1.12	a*
	2	105	3.68	0.75	a*; b**
	3	82	3.28	0.89	b**
Frequency of drug use	1	11	0.65	0.88	
	2	159	0.13	0.39	a**
	3	104	0.44	0.99	a**
Frequency of tobacco use	1	11	1.91	1.76	
	2	159	0.89	1.40	a**
	3	104	1.57	1.70	a**
Frequency of alcohol use	1	11	1.73	1.35	
	2	159	1.85	1.27	a*
	3	104	2.23	1.16	a*
Frequency of unprotected sexual intercourse	1	7	1.86	1.57	a*
	2	104	0.61	1.23	a*
	3	65	0.85	1.43	

Note. The same letters indicate a significant difference between the means.
* $p \leq .05$.
** $p \leq .01$.

$p < .01$; $\eta^2 = 0.036$). In interpreting these results, it is worth noting that the effect sizes were small. Compared with problematic (Class 1) and anxious/antisocial (Class 3) individuals, well-adjusted individuals (Class 2) showed lower anxiety and depression, and higher satisfaction for social relations (see Table 9). In addition, these individuals reported less frequent use of drug, tobacco, and alcohol than anxious/antisocial individuals, and less frequent unprotected sexual intercourse than problematic individuals. Problematic and anxious/antisocial individuals did not differ in any of these variables. It is worth noting that some nonsignificant differences between Class 1 and the other two classes might be due to the small size of Class 1.

3.3. Discussion

This study aimed to validate the abbreviated form of the EPQ-R developed in Study 1 on a new data sample. The results supported the construct validity of the instrument: The four-factor structure was confirmed, and all convergent validity coefficients were coherent with expectations.

Three latent patterns were identified through LCA. The first pattern consisted of individuals with high values of P and N. The second pattern consisted of individuals with low levels of P and N, and high scores of L. With respect to this latter dimension, it should be noted that L scores may be interpreted as the tendency of individuals to provide an inflated image of themselves, but also as the tendency of being conformist (Eysenck & Eysenck, 1975; Massey, 1980; McCrae & Costa, 1983; Uziel, 2010, 2014). Finally, the third pattern consisted of individuals characterized by relatively high levels of P and N. The individuals belonging to these three patterns have been labelled as problematic, well-adjusted, and anxious/antisocial, respectively. Well-adjusted individuals were found to differ from problematic and anxious/antisocial individuals in indicators of psychosocial functioning, such as anxiety and depression, satisfaction for social relations, frequency of substance use and sexual risk behaviors.

4. Final remarks

This work aimed to develop a new abbreviated form of the EPQ-R. To this purpose, two studies were carried out. In Study 1, a MIRT model was estimated on the 100 items of the full-length version of the questionnaire to identify 24 items (6 per scale) which were simple-structured and unbiased, well-fitting, well-discriminating, and well-distributed along the latent traits continua. These items were used to compose a new abbreviated form of the questionnaire. In Study 2, the questionnaire was validated on a new data sample. The new abbreviated form of the EPQ-R outperformed the old one in reliability and approximation of measures obtained with the full-length test. Moreover, the four-factor structure of the instrument was confirmed, and all convergent validity coefficients were in line with expectations. These results provide further evidence about the usefulness of IRT approaches for the development of valid and reliable measurement instruments.

Despite notable improvements, reliability of the P scale remains rather low. This result, however, was expected. This scale has been recognized in literature as the most problematic and controversial of the instrument (e.g., Bishop, 1977; Block, 1977; Claridge, 1981), and previous studies conducted both in Italy (e.g., Colledani, Anselmi, & Robusto, 2018; Colledani, Robusto, & Anselmi, 2018) and in other countries (e.g., Forrest et al., 2000; Shevlin et al., 2002) have highlighted its psychometric weaknesses. Future studies may be devoted to develop a new pool of items more suitable for the assessment of this trait.

In this study, three distinct patterns of the four PEN-L traits have been identified. At the best of our knowledge, this is the first study that investigated the latent patterns underlying the PEN-L traits. The results seem promising. The emerged patterns are rather well-defined and identifiable. Moreover, they differ with respect to indicators of psychosocial functioning. Results of this study suggest that individuals characterized by particular patterns of PEN-L traits might be particularly exposed to risks (e.g., contraction of sexually transmitted disease, drug addiction). Future studies should further explore these patterns and their existence in other cross-cultural contexts.

References

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education, 7*, 255–278. https://doi.org/10.1207/s15324818ame0704_1.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.). *Proceedings of the 2nd international symposium on information* (pp. 267–281). Budapest: Akademiai Kiado.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Anselmi, P., Vianello, M., Voci, A., & Robusto, E. (2013). Implicit sexual attitude of heterosexual, gay and bisexual individuals: Disentangling the contribution of specific associations to the overall measure. *PLoS One, 8*(11), e78990. https://doi.org/10.1371/journal.pone.0078990.

Anselmi, P., Vidotto, G., Bettinardi, O., & Bertolotti, G. (2015). Measurement of change in health status with Rasch models. *Health and Quality of Life Outcomes, 13*, 16. https://doi.org/10.1186/s12955-014-0197-x.

Balsamo, M., Giampaglia, G., & Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatric Disease and Treatment, 10*, 153–165. https://doi.org/10.2147/NDT.S53425.

Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences, 34*, 645–664. https://doi.org/10.1016/S0191-8869(02)00051-X.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238.

van den Berg, S. M., Paap, M. C., Derks, E. M., & Outcome of Psychosis (GROUP) investigators (2013). Using multidimensional modeling to combine self-report symptoms with clinical judgment of schizotypy. *Psychiatry Research, 206*(1), 75–80. https://doi.org/10.1016/j.psychres.2012.09.015.

Bishop, D. V. (1977). The P scale and psychosis. *Journal of Abnormal Psychology, 86*, 127–134. https://doi.org/10.1037/0021-843X.86.2.127.

Block, J. (1977). P scale and psychosis: Continued concerns. *Journal of Abnormal Psychology, 86*, 431–434. https://doi.org/10.1037/0021-843X.86.4.431.

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice, 16*, 21–33. https://doi.org/10.1111/j.1745-3992.1997.tb00605.x.

Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., & de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. *Quality & Quantity, 47*, 2341–2360. https://doi.org/10.1007/s11135-012-9684-5.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.). *Testing structural equation models* (pp. 136–162). Newbury Park, CA: SAGE.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195–212. https://doi.org/10.1007/BF01246098.

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K.-M., Falk, C. F., ... Oguzhan, O. (2018). Package "mirt" (version 1.29). Retrieved from https://cran.r-project.org/web/packages/mirt/mirt.pdf.

Cheng, Y. Y., Wang, W. C., & Ho, Y. H. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: A statistical solution for the bandwidth—Fidelity dilemma. *Educational and Psychological Measurement, 69*, 369–388. https://doi.org/10.1177/0013164408323241.

Chiorri, C., Marsh, H. W., Ubbiali, A., & Donati, D. (2016). Testing the factor structure and measurement invariance across gender of the Big Five Inventory through exploratory structural equation modeling. *Journal of Personality Assessment, 98*, 88–99. https://doi.org/10.1080/00223891.2015.1035381.

Claridge, G. S. (1981). Psychoticism. In R. Lynn (Ed.). *Dimensions of personality* (pp. 79–110). Oxford, UK: Pergamon Press.

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

Colledani, D. (2018). Psychometric properties and gender invariance for the Dickman Impulsivity Inventory. *Testing, Psychometrics, Methodology in Applied Psychology, 25*, 49–61. https://doi.org/10.4473/TPM25.1.3.

Colledani, D., Anselmi, P., & Robusto, E. (2018). Using item response theory for the development of a new short form of the Eysenck Personality Questionnaire-Revised. *Frontiers in Psychology, 9*, 1834. https://doi.org/10.3389/fpsyg.2018.01834.

Colledani, D., Anselmi, P., & Robusto, E. (2019). Using multidimensional item response theory to develop an abbreviated form of the Italian version of Eysenck's IVE questionnaire. *Personality and Individual Differences, 142*, 45–52. https://doi.org/10.1016/j.paid.2019.01.032.

Colledani, D., Robusto, E., & Anselmi, P. (2018). Development of a new abbreviated form of the Junior Eysenck Personality Questionnaire-Revised. *Personality and Individual Differences, 120*, 159–165. https://doi.org/10.1016/j.paid.2017.08.037.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences.* Hoboken, NJ: John Wiley & Sons.

Corulla, W. J. (1990). A revised version of the psychoticism scale for children. *Personality and Individual Differences, 11*(1), 65–76. https://doi.org/10.1016/0191-8869(90)90169-R.

Da Dalt, L., Anselmi, P., Bressan, S., Carraro, S., Baraldi, E., Robusto, E., & Perilongo, G. (2013). A short questionnaire to assess pediatric resident's competencies: The validation process. *Italian Journal of Pediatrics, 39*, 41. https://doi.org/10.1186/1824-7288-39-41.

Da Dalt, L., Anselmi, P., Furlan, S., Carraro, S., Baraldi, E., Robusto, E., & Perilongo, G. (2015). Validating a set of tools designed to assess the perceived quality of training of pediatric residency programs. *Italian Journal of Pediatrics, 41*, 2. https://doi.org/10.1186/s13052-014-0106-2.

Da Dalt, L., Anselmi, P., Furlan, S., Carraro, S., Baraldi, E., Robusto, E., & Perilongo, G. (2017). An evaluation system for postgraduate pediatric residency programs: Report of a 3-year experience. *European Journal of Pediatrics, 176*, 1279–1283. https://doi.org/10.1007/s00431-017-2967-z.

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. https://doi.org/10.1016/j.paid.2010.12.032.

Dazzi, C., Pedrabissi, L., & Santinello, M. (2004). *EPQ-R Eysenck personality questionnaire revised.* Florence, Italy: Giunti OS Organizzazioni Speciali.

Draycott, S. G., & Kline, P. (1995). The big three or the big five—The EPQ-R vs the NEO-PI: A research note, replication and elaboration. *Personality and Individual Differences, 18*, 801–804. https://doi.org/10.1016/0191-8869(95)00010-4.

Escorial, S., & Navas, M. J. (2007). Analysis of the gender variable in the Eysenck Personality Questionnaire–Revised scales using differential item functioning techniques. *Educational and Psychological Measurement, 67*, 990–1001. https://doi.org/10.1177/0013164406299108.

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

Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scale (adults).* London, UK: Hodder and Stoughton.

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6*, 21–29. https://doi.org/10.1016/0191-8869(85)90026-1.

Eysenck, S. B., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences, 6*, 613–619. https://doi.org/10.1016/0191-8869(85)90011-X.

Ferguson, S. L., & Hull, D. M. (2018). Personality profiles: Using latent profile analysis to model personality typologies. *Personality and Individual Differences, 122*, 177–183. https://doi.org/10.1016/j.paid.2017.10.029.

Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis–based models. *Applied Psychological Measurement, 34*, 10–26. https://doi.org/10.1177/0146621609336112.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491. https://doi.org/10.1037/1082-989X.9.4.466.

Forrest, S., Lewis, C. A., & Shevlin, M. (2000). Examining the factor structure and differential functioning of the Eysenck Personality Questionnaire Revised-Abbreviated. *Personality and Individual Differences, 29*(3), 579–588. https://doi.org/10.1016/S0191-8869(99)00220-2.

Francis, L. J. (1996). The relationship between Eysenck's personality factors and attitude towards substance use among 13–15-year-olds. *Personality and Individual Differences, 21*, 633–640. https://doi.org/10.1016/0191-8869(96)00125-0.

Francis, L. J., Brown, L. B., & Philipchalk, R. (1992). The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the USA and Australia. *Personality and Individual Differences, 13*, 443–449. https://doi.org/10.1016/0191-8869(92)90073-X.

Francis, L. J., & Pearson, P. R. (1988). Religiosity and the short-scale EPQ-R indices of E, N and L, compared with the JEPI, JEPQ and EPQ. *Personality and Individual Differences, 9*, 653–657. https://doi.org/10.1016/0191-8869(88)90162-6.

Haberman, S. J., von Davier, M., & Lee, Y.-H. (2008). *Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions (RR-08-45).* Educational Testing Service.

Heaven, P. C., Ciarrochi, J., Leeson, P., & Barkus, E. (2013). Agreeableness, conscientiousness, and psychoticism: Distinctive influences of three personality dimensions in adolescence. *British Journal of Psychology, 104*(4), 481–494. https://doi.org/10.1111/bjop.12002.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. https://doi.org/10.1080/10705519909540118.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.). *Handbook of personality: Theory and research* (pp. 114–158). New York, NY: Guilford Press.

Karanci, N., Dirik, G., & Yorulmaz, O. (2007). Reliability and validity studies of Turkish translation of Eysenck personality questionnaire revised-abbreviated. *Türk Psikiyatri Dergisi, 18*, 254–261.

Kinnunen, M. L., Metsäpelto, R. L., Feldt, T., Kokko, K., Tolvanen, A., Kinnunen, U., ... Pulkkinen, L. (2012). Personality profiles and health: Longitudinal evidence among Finnish adults. *Scandinavian Journal of Psychology, 53*(6), 512–522. https://doi.org/10.1111/j.1467-9450.2012.00969.x.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology, 23*(1), 84–86. https://doi.org/10.1002/1097-4679(196701)23:1<84::AID-JCLP2270230123>3.0.CO;2-2.

Lo, Y. T., Mendell, N. R., & Rubin, B. D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778. https://doi.org/10.1093/biomet/88.3.767.

Luecht, R. M., & Miller, T. R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement, 16*(3), 279–293. https://doi.org/10.1177/014662169201600308.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341. https://doi.org/10.1207/s15328007sem1103_2.

Massey, A. (1980). The Eysenck Personality Inventory lie scale: Lack of insight or…? *The Irish Journal of Psychology, 4*(3), 172–174.

McCrae, R. R., & Costa, P. T., Jr. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6*, 587–597. https://doi.org/10.1016/0191-8869(85)90008-X.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882–888. https://doi.org/10.1037/0022-006X.51.6.882.

Merz, E. L., & Roesch, S. C. (2011). A latent profile analysis of the Five Factor Model of personality: Modeling trait interactions. *Personality and Individual Differences, 51*, 915–919. https://doi.org/10.1016/j.paid.2011.07.022.

Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika, 74*, 273–296. https://doi.org/10.1007/s11336-008-9097-5.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus version 7 user's guide.* Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569. https://doi.org/10.1080/10705510701575396.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*, 50–64. https://doi.org/10.1177/01466216000241003.

Parr, A. D., Lanza, S. T., & Bernthal, P. (2016). Personality profiles of effective leadership performance in assessment centers. *Human Performance, 29*, 143–157. https://doi.org/10.1080/08959285.2016.1157596.

Pedrabissi, L., & Santinello, M. (1989). *Inventario per L'ansia di "Stato" e di "Tratto": Nuova Versione Italiana Dello Stai Forma Y: Manuale.* Firenze, IT: Organizzazioni Speciali.

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health, 18*, 25–34. https://doi.org/10.1016/j.jval.2014.10.005.

Reckase, M. D. (2009). *Multidimensional item response theory: Statistics for social and behavioral sciences.* New York: Springer.

Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment, 7*, 347–364. https://doi.org/10.1177/107319110000700404.

Saggino, A. (2000). The big three or the big five? A replication study. *Personality and Individual Differences, 28*, 879–886. https://doi.org/10.1016/S0191-8869(99)00146-4.

Scholte, R. H., & De Bruyn, E. E. (2004). Comparison of the Giant Three and the Big Five in early adolescents. *Personality and Individual Differences, 36*, 1353–1371. https://doi.org/10.1016/S0191-8869(03)00234-4.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343. https://doi.org/10.1007/bf02294360.

Shevlin, M., Bailey, F., & Adamson, G. (2002). Examining the factor structure and sources of differential functioning of the Eysenck Personality Questionnaire Revised—Abbreviated. *Personality and Individual Differences, 32*(3), 479–487. https://doi.org/10.1016/S0191-8869(01)00049-6.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice, 30*(3), 29–40. https://doi.org/10.1111/j.1745-3992.2011.00208.x.

Sotgiu, I., Anselmi, P., & Meneghini, A. M. (2019). Investigating the psychometric properties of the Questionnaire for Eudaimonic Well-Being: A Rasch analysis. *TPM–testing, psychometrics, methodology in applied psychology* (in press).

Spence, R., Owens, M., & Goodyer, I. (2012). Item response theory and validity of the NEO-FII in adolescents. *Personality and Individual Differences, 53*, 801–807. https://doi.org/10.1016/j.paid.2012.06.00.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory STAI.* Palo Alto, CA: Consulting Psychologists Press.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Patient Health Questionnaire Primary Care Study Group (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA, 282*, 1737–1744. https://doi.org/10.1001/jama.282.18.1737.

Thissen, D., & Steinberg, L. (2009). Item response theory. In R. Millsap, & A. Maydeu-Olivares (Eds.). *The Sage handbook of quantitative methods in psychology* (pp. 148–177). London: Sage.

Ubbiali, A., Chiorri, C., & Hampton, P. (2013). Italian Big Five Inventory. Psychometric properties of the Italian adaptation of the Big Five Inventory (BFI). *BPA-Applied Psychology Bulletin (Bollettino di Psicologia Applicata), 59*(266), 37–48.

Uziel, L. (2010). Rethinking social desirability scales from impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*(3), 243–262. https://doi.org/10.1177/1745691610369465.

Uziel, L. (2014). Impression management ("lie") scales are associated with interpersonally oriented self-control, not other-deception. *Journal of Personality, 82*, 200–212. https://doi.org/10.1111/jopy.12045.

Vidotto, G., Anselmi, P., Filipponi, L., Tommasi, M., & Saggino, A. (2018). Using overt and covert items in self-report personality tests: Susceptibility to faking and identifiability of possible fakers. *Frontiers in Psychology, 9*, 1100. https://doi.org/10.3389/fpsyg.2018.01100.

Yao, L., & Boughton, K. (2009). Multidimensional linking for tests with mixed item types. *Journal of Educational Measurement, 46*, 177–197. https://doi.org/10.1111/j.1745-3984.2009.00076.x.

Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Doctoral dissertation). Retrieved from https://www.statmodel.com/download/Yudissertation.pdf.

Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica, 29*, 18. https://doi.org/10.1186/s41155-016-0040-x.

Zhang, J., Bray, B. C., Zhang, M., & Lanza, S. T. (2015). Personality profiles and frequent heavy drinking in young adulthood. *Personality and Individual Differences, 80*, 18–21. https://doi.org/10.1016/j.paid.2015.01.054.