



# Using Item Response Theory for the Development of a New Short Form of the Eysenck Personality Questionnaire-Revised

*Daiana Colledani\**, Pasquale Anselmi and Egidio Robusto

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

## OPEN ACCESS

### Edited by:

Marco Innamorati,  
Università Europea di Roma, Italy

### Reviewed by:

Davide Marengo,  
Università degli Studi di Torino, Italy  
Pietro Cipresso,  
Istituto Auxologico Italiano (IRCCS),  
Italy

### \*Correspondence:

Daiana Colledani  
daianacolledani@gmail.com

### Specialty section:

This article was submitted to  
Quantitative Psychology and  
Measurement,  
a section of the journal  
Frontiers in Psychology

**Received:** 15 March 2018

**Accepted:** 07 September 2018

**Published:** 02 October 2018

### Citation:

Colledani D, Anselmi P and Robusto E  
(2018) Using Item Response Theory  
for the Development of a New Short  
Form of the Eysenck Personality  
Questionnaire-Revised.  
*Front. Psychol.* 9:1834.  
doi: 10.3389/fpsyg.2018.01834

The present work aims at developing a new version of the short form of the Eysenck Personality Questionnaire-Revised, which includes Psychoticism, Extraversion, Neuroticism, and Lie scales (48 items, 12 per scale). The work consists of two studies. In the first one, an item response theory model was estimated on the responses of 590 individuals to the full-length version of the questionnaire (100 items). The analyses allowed the selection of 48 items well discriminating and distributed along the latent continuum of each trait, and without misfit and differential item functioning. In the second study, the functioning of the new form of the questionnaire was evaluated in a different sample of 300 individuals. Results of the two studies show that reliability of the four scales is better than, or equal to that of the original forms. The new version outperforms the original one in approximating scores of the full-length questionnaire. Moreover, convergent validity coefficients and relations with clinical constructs were consistent with literature.

**Keywords:** short Eysenck personality questionnaire-revised, item response theory, 2PL, ESEM, DIF

## INTRODUCTION

In the view of Eysenck (see Eysenck and Eysenck, 1975, 1991), the structure of personality may be effectively described by three main traits: psychoticism (P), extraversion (E), and neuroticism (N). These dimensions are also known as the “Giants Three” and represent basic, independent, and biologically founded traits. They characterize all subjects, with varying degrees, and allow for effectively describing behavioral, emotional, and individual differences among adults and young people. According to the authors, PEN traits do not represent pathological dimensions in themselves, but could lead to the development of abnormal conditions only in particular situations (Eysenck and Eysenck, 1991). In this perspective, neurosis and psychosis should be conceived as pathological exaggerations of the underlying traits of neuroticism and psychoticism (Eysenck and Eysenck, 1991; Mor, 2010).

Extraversion and neuroticism have been the first two dimensions included in the Eysenck’s model and were conceptualized as orthogonal continua (Eysenck and Eysenck, 1964, 1991). The neuroticism dimension describes a trait opposed to emotional stability, and defines the degree to which a person is predisposed to experience negative affect (Eysenck and Eysenck, 1964, 1991; Mor, 2010). Individuals with high levels of this trait tend to be worried, apprehensive, moody, fed-up, and irritable (Eysenck and Eysenck, 1991; Eysenck and Barrett, 2013). Extraversion is the second dimension included in the model and depicts sociable, carefree,

friendly, convivial, easygoing, and impulsive individuals. This trait is opposed to introversion which, in contrast, defines individuals introspective, quiet, serious, and reserved (Eysenck and Eysenck, 1975, 1991; Eysenck and Barrett, 2013). The third dimension included in the Eysenck's model has been psychoticism, or toughmindedness. The typical toughminded is an individual hostile, aggressive, untrusting, cold, unemotional, rude, lacking in human feelings, and unfriendly. On the opposite pole of the continuum, there are individuals with well-adjusted personality, agreeable, empathic, tolerant, conscientious, open-minded, friendly, and warm (Eysenck and Eysenck, 1975, 1991; Eysenck and Barrett, 2013).

Over the years, a series of instruments has been developed for the assessment of PEN traits on both young and adult people (e.g., Eysenck and Eysenck, 1964, 1975; Eysenck et al., 1985). These instruments also included a Lie (L) scale, which measures dissimulation and the tendency to deceive (Eysenck and Eysenck, 1964). Several contributions have been offered for the refinement of the psychometric properties of Eysenck's questionnaires, as well as for the development of brief versions (Eysenck et al., 1985; Francis and Pearson, 1988; Corulla, 1990; Francis et al., 1992; Francis, 1996). The psychometric properties and factor structure of all these instruments have been investigated in cross cultural research (e.g., Hosokawa and Ohyama, 1993; Maltby and Talley, 1998; Forrest et al., 2000; Qian et al., 2000; Scholte and De Bruyn, 2001; Aluja et al., 2003; Alexopoulos and Kalaitzidis, 2004; Dazzi et al., 2004; Francis et al., 2006; Karanci et al., 2006; Tiwari et al., 2009; Picconi et al., 2018). Unidimensionality of N and L scales has been widely supported in literature (e.g., Lajunen and Scherler, 1999; Ferrando, 2001; Ferrando and Chico, 2001; Ferrando and Anguiano-Carrasco, 2009; Dazzi, 2011). Contrasting results have been found concerning E scale: There are several studies supporting the unidimensionality of this scale (e.g., Rocklin and Revelle, 1981; Ferrando and Chico, 2001; Dazzi, 2011), but there is also some evidence suggesting the presence of two dimensions (Eysenck and Eysenck, 1963; Vidotto et al., 2008). Finally, there is large agreement in the literature that P scale comprises different facets (e.g., Howarth, 1986; Roger and Morris, 1991), which nevertheless contribute to a unique dimension (Chico and Ferrando, 1995; Dazzi, 2011).

Eysenck's instruments have been extensively employed for clinical, forensic, educational, and organizational purposes (e.g., Nyborg, 1997; Judge et al., 2000; Wood and Newton, 2003; Laidra et al., 2007; Smillie et al., 2009; Almiro et al., 2016), and all scales showed significant relations with a variety of psychologically and clinically relevant constructs and behaviors. Research, for instance, suggests that individuals with high levels of neuroticism may experience symptoms of anxiety and depression (e.g., Eysenck, 1991; Saklofske et al., 1995; del Barrio et al., 1997; Dazzi et al., 2004; Jylhä and Isometsä, 2006), and may also be more likely exposed to stress and health problems (e.g., Denney and Frisch, 1981; Huang et al., 2015; Bergomi et al., 2017). In contrast, extraversion appears to be mainly linked to adaptive social behavior, mental well-being, happiness, and life satisfaction (e.g., Lu, 1995; Mor, 2010; Gale et al., 2013). Moreover, this trait has been found to be negatively related to symptoms of anxiety and depression, to self-reported mental disorder and to health

care use for psychiatric reasons (e.g., del Barrio et al., 1997; Jylhä and Isometsä, 2006). Finally, psychoticism has been often cited in relation to inappropriate social behaviors, such as unsafe sexual habits, heavy drinking, criminal behavior, dysfunctional impulsivity, gambling, and drug abuse (e.g., Barnes et al., 1984; Blaszczyński et al., 1985; Bogaert, 1993; Lodhi and Thakur, 1993; Francis, 1996; Conrad et al., 1997; Grau and Ortet, 1999; Hoyle et al., 2000; Chico et al., 2003; Heaven et al., 2004; Gudgeon et al., 2005; Colledani, 2018).

The short form of the Eysenck Personality Questionnaire-Revised (EPQ-R; Eysenck et al., 1985; Eysenck and Eysenck, 1991) includes 48 items (out of 100 of the EPQ-R), 12 per each of the four dimensions. This version of the instrument has been translated in several languages and is widely used, across different countries, for scientific and clinical purposes (Hosokawa and Ohyama, 1993; Aluja et al., 2003; Alexopoulos and Kalaitzidis, 2004; Dazzi et al., 2004; Francis et al., 2006; Tiwari et al., 2009; Sanavio et al., 2013). However, it suffers from the same drawbacks of the full-length version. In particular, P scale exhibited poor reliability with a restricted range of scores and a strong positive skewness (Bishop, 1977; Block, 1977; Claridge, 1981; Hosokawa and Ohyama, 1993; Katz and Francis, 2000; Alexopoulos and Kalaitzidis, 2004). In addition, several items showed differential item functioning (DIF) across gender (Eysenck et al., 1985; Eysenck and Eysenck, 1991; Lynn and Martin, 1997; Forrest et al., 2000; Karanci et al., 2006; Escorial and Navas, 2007), which makes the comparison between groups questionable.

A better selection of the items from the full-length version of the instrument could allow for reducing some of the aforementioned drawbacks. The present work aims at developing a new version of the short form of the EPQ-R with improved psychometric properties.

Item response theory (IRT; Bock, 1997; Thissen and Steinberg, 2009) is one of the most promising approaches to this aim. There are several successful applications of IRT for the development and validation of measurement scales (see, Da Dalt et al., 2013, 2015; Balsamo et al., 2014; Anselmi et al., 2015; Zanon et al., 2016; Sotgiu et al., 2018). Moreover, compared with classical test theory, IRT was found to provide more diagnostic information useful for the development of brief scales (Spence et al., 2012; Bortolotti et al., 2013; Petrillo et al., 2015). IRT allows for identifying the items that are best at discriminating different levels of the latent trait of interest, while ensuring that the entire trait continuum is covered. Selecting these items can result in a brief version of the scale that produces scores very similar to those obtained with the full-length scale and has the same external validity (i.e., the same correlations with other constructs; Reise and Henson, 2000; Spence et al., 2012). Moreover, IRT allows for detecting items that are unclear, ambiguous, or which exhibit DIF. These items should be not included in the brief scale. Despite advantages offered by IRT, only a few studies employed this approach for the refinement of Eysenck's instruments (e.g., Ferrando, 2001; Ferrando and Chico, 2001; Escorial and Navas, 2007; Majj-de Meij et al., 2008). Recently, Colledani et al. (2018) used IRT for developing a new version of the abbreviated form of the Junior EPQ-R (6 items per scale). The new version outperformed the original one on several aspects.

This work includes two main studies. In Study 1, a series of analyses were performed on the responses to the full-length version of the EPQ-R in order to select the 48 items (12 per each scale) with the best psychometric properties. In Study 2, the functioning of the new short form was tested in a new data sample. Reliability, validity and factor structure were examined. Relationships of the new scales with social desirability, the dimensions of the Five Factor Model (FFM), and clinically relevant constructs were verified.

## STUDY 1

### Participants

A total of 590 participants took part in the study (mean age = 36.69 years,  $SD = 14.16$ ; from 18 to 75 years; 55.8% females). They were recruited from different Italian regions through convenience sampling. All participants were native Italian speakers and completed the questionnaire anonymously and voluntarily. All standards for research with human subjects were respected. Written informed consent was obtained from the participants. The project has been approved, now as later, by the Ethical Committee for the Psychological Research of the University of Padova since a prospective ethics approval was not required at the time when the research was conducted (Protocol n. 2622).

### Instruments

The participants were presented with the Italian version of the EPQ-R (Dazzi et al., 2004; Dazzi, 2011). The instrument 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?”). Administration of the questionnaire was individual and paper-and-pencil.

The Italian version of the questionnaire has good reliability and the four-factor structure was confirmed ( $\alpha = 0.67, 0.78, 0.85$ , and  $0.75$  for P, E, N, and L scales, respectively; Dazzi et al., 2004; Dazzi, 2011). The reliability found in the current sample ( $\alpha = 0.60, 0.79, 0.85$ , and  $0.77$  for P, E, N, and L scales) is in line with literature.

Studies in the Italian context aimed also to test the factor structure and the psychometric characteristics of the short version of the instrument (Dazzi et al., 2004). Consistently with cross-cultural findings, results supported the four-factor structure of the instrument and showed reliability coefficients satisfactory for E, N, and L scales, while lower for P ( $\alpha = 0.37, 0.77, 0.83$ , and  $0.70$  for P, E, N, and L, respectively; Dazzi et al., 2004). The reliability found in the current sample ( $\alpha = 0.40, 0.73, 0.83$ , and  $0.73$  for P, E, N, and L scales) is in line with literature.

### Analysis Strategy

The two-parameter logistic (2PL) model (see Thissen and Steinberg, 2009) was separately estimated on the responses to each of the four scales of the questionnaire. This model describes

the probability that a subject endorses a certain item as a function of the latent trait level of the subject (parameter  $\theta$ ), the “endorsability” level of the item (i.e., the ease of providing a “yes” response to that item; parameter  $\epsilon$ ), and the capability of the item in differentiating subjects with different trait levels (parameter  $\delta$ ). In the case of the P scale, for instance, the greater the value of parameter  $\theta$ , the greater the level of psychoticism of the subject; the greater the value of parameter  $\epsilon$ , the greater the ease of responding “yes” to the item (i.e., of providing a response that is indicative of the presence of psychoticism); the greater the value of parameter  $\delta$ , the greater the capability of the item in differentiating between subjects with different levels of psychoticism. All the analyses were run using the packages “difR” (Magis et al., 2016) and “ltm” (Rizopoulos, 2012) for the statistical environment R (R Core Team, 2016).

The 2PL assumes unidimensionality of the scales. Confirmatory factor analyses were run on the data of each of the four scales (for a reasonable fit,  $CFI \geq 0.90$ ,  $RMSEA < 0.08$ ; see Hu and Bentler, 1999; Marsh et al., 2004; Brown, 2006). These analyses confirmed the unidimensionality of N [ $\chi^2_{(252)} = 1046.791, p \leq 0.001$ ;  $RMSEA = 0.073$ ;  $CFI = 0.919$ ] and L [ $\chi^2_{(189)} = 532.901, p \leq 0.001$ ;  $RMSEA = 0.056$ ;  $CFI = 0.900$ ]. Fit indices of E scale were close to acceptance [ $\chi^2_{(230)} = 808.417, p \leq 0.001$ ;  $RMSEA = 0.065$ ;  $CFI = 0.890$ ]. The unidimensional model did not fit the data of P scale [ $\chi^2_{(464)} = 1841.233, p \leq 0.001$ ;  $RMSEA = 0.071$ ;  $CFI = 0.467$ ]. An exploratory factor analysis on this scale suggests a four-factor solution with 7 items out of 32 exhibiting cross-loadings. In line with literature (e.g., Howarth, 1986; Roger and Morris, 1991; Chico and Ferrando, 1995; Dazzi, 2011), this result confirms that P scale defines a complex and multifaceted construct.

### Item Selection for the New Short Scales

DIF and item fit statistics were used to identify the items with the poorest psychometric properties that were not included in the new short scales.

Three item fit statistics were used: *infit*, *outfit* (Wright and Masters, 1982), and the index suggested by Bock (1972). *Infit* and *outfit* are two  $\chi^2$ -based statistics, the former being effective in detecting unexpected responses to items close to a subject's trait level, the latter being effective in detecting unexpected responses to items far from the subject's trait level. In this work, items with *infit* and/or *outfit* higher than 1.4 (Wright and Linacre, 1994) were considered misfitting and not included in the new short scales. The index suggested by Bock involves grouping subjects into  $n$  categories on the basis of their latent trait level, and observed and expected proportions of subjects endorsing the item for each group are compared (Bock, 1972; Reise, 1990). In this work, subjects were grouped into four categories and the items which displayed a medium ( $0.3 \leq \Phi < 0.5$ ) to large ( $\Phi \geq 0.5$ ) effect size (Cohen, 1988) were not selected for inclusion in the new questionnaire.

Items exhibiting gender DIF were also excluded from the new questionnaire. Both uniform and non-uniform DIF were considered. The former is a systematic bias expressing a different probability of endorsing an item for the members of a specific group. The latter is a non-systematic bias which varies with the

latent trait level. Females were used as reference group. Effect sizes of uniform and non-uniform DIF were evaluated by the  $R^2$  difference test (Nagelkerke, 1991; Gómez-Benito et al., 2009), with values higher than 0.035 denoting moderate DIF and values higher than 0.07 denoting strong DIF (Jodoin and Gierl, 2001; Magis et al., 2016).

Parameters  $\epsilon$  and  $\delta$  were examined to select, among the remaining items, those that allow for covering the entire trait continuum and with the greatest discrimination level.

### Assessment of the Psychometric Characteristics of the New Short Scales

Reliability and validity of the newly developed PEN-L scales were evaluated and compared with those of the original short

scales. Reliability was evaluated through Cronbach's  $\alpha$  and test information function (TIF). TIF tells us how well the test measures the latent trait levels over the entire range of interest (Baker, 2001; Petrillo et al., 2015). The larger the value of TIF, the greater the accuracy with which the latent trait levels are measured. TIF depends on the latent trait range under consideration and on the number of items in the test (Baker, 2001). In this work, the old and new short scales had the same length (12 items), and TIF was defined on the same range of latent trait levels ( $-5$  to  $5$ ). Validity was evaluated using a bias index and the correlation between scores obtained with full-length and short scales. The bias index was computed as the average difference (in absolute terms) between the parameters  $\theta$  estimated on the full-length scales and those estimated on the short scales.

**TABLE 1** | Easiness ( $\epsilon$ ) and discrimination ( $\delta$ ) parameters for the 32 items of the Psychoticism scale.

New form	Original form	Item n.	Text	$\epsilon$	$\delta$
		99	Would you feel very sorry for an animal caught in a trap?	-11.873	-0.121
		56	Do most things taste the same to you?	-10.789	-0.195
	✓	79	Do you try not to be rude to people?	-7.722	-0.354
		95	Do people tell you a lot of lies?	-2.538	-0.200
✓	✓	29	Do you prefer to go your own way rather than act by the rules?	-0.855	0.939
		9	Do you give money to charities?	-0.395	0.373
✓	✓	88	Is it better to follow society's rules than go your own way?	-0.045	1.247
✓		42	Have you often gone against your parents' wishes?	0.466	0.703
	✓	5	Do you take much notice of what people think?	0.502	0.150
	✓	75	Do you think people spend too much time safeguarding their future with savings and insurance?	0.713	0.313
✓		18	Should people always respect the law?	0.750	1.260
✓		64	Do you like to arrive at appointments in plenty of time?	0.953	0.846
✓		21	Are good manners very important?	0.993	1.580
✓	✓	41	Do good manners and cleanliness matter much to you?	1.439	1.763
		2	Do you stop to think things over before doing anything?	1.448	0.569
✓		81	Do you generally 'look before you leap'?	1.807	0.708
✓		96	Do you believe one has special duties to one's family?	2.404	1.249
✓	✓	25	Would you take drugs which may have strange or dangerous effects?	2.725	1.138
	✓	48	Do you think marriage is old-fashioned and should be done away with?	3.343	0.552
		34	Do you have enemies who want to harm you?	3.734	0.367
✓		30	Do you enjoy hurting people you love?	3.775	0.910
	✓	59	Does it worry you if you know there are mistakes in your work?	4.024	0.538
	✓	7	Would being in debt worry you?	4.028	0.446
		73	Are there several people who keep trying to avoid you?	4.119	0.596
		68	Is (or was) your mother a good woman?	4.774	0.527
		85	Can you on the whole trust people to tell the truth?	4.986	-0.128
		12	Would it upset you a lot to see a child or an animal suffer?	5.307	0.750
	✓	54	Do you enjoy co-operating with others?	5.517	0.446
		14	Would it upset you a lot to see a child or an animal suffer?	5.623	-0.044
	✓	91	Would you like other people to be afraid of you?	6.036	0.242
✓		37	Do you have many friends?	6.638	0.515
		50	Are you more easy-going about right and wrong than most people?	8.580	-0.089

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."



Low biases suggest that the latent trait estimates obtained with the short scales approximate those of the full-length versions. In addition, the correlations between scores obtained with the full-length and short scales were computed and corrected for common items using the Levy's (1967) method.

## Results

Three of the 32 items of P scale exhibited uniform and non-uniform gender DIF of moderate (Items 68 and 91) or strong (Item 12) size. Fit statistics were adequate for all the items. From the remaining 29 items, 12 were selected taking into account their parameters  $\epsilon$  and  $\delta$ . This resulted in a new short scale, that differed from the original one for eight items (see **Table 1**). Specifically, Item 91 was changed because it showed uniform and non-uniform gender DIF of moderate size. These modifications allowed for obtaining a new scale with increased reliability ( $\alpha$  increased from 0.40 to 0.62; TIF increased from 8.13 to 12.86) and with scores that better approximate those obtained with the full-length scale (bias decreased from 0.37 to 0.18, corrected correlation increased from 0.47 to 0.52). It is worth noting that Cronbach's  $\alpha$  of the new 12-item scale (0.62) largely resembles that of the full 32-item scale (0.60).

Regarding the 23 items of E scale, only Item 55 exhibited uniform gender DIF of moderate size and no item showed misfit. Selecting 12 items upon the basis of their parameters  $\epsilon$  and  $\delta$ , we obtained a new E scale that differed from the original one for three items (see **Table 2**). The differences in reliability and validity of the new and original scales were small in size,

nevertheless in favor of the new version ( $\alpha$  increased from 0.73 to 0.75; TIF increased from 16.62 to 16.83; bias decreased from 0.21 to 0.19; corrected correlation increased from 0.74 to 0.77).

Concerning N and L scales, no one item exhibited gender DIF or misfit. Therefore, items were selected considering their  $\epsilon$  and  $\delta$  parameters. For both scales, the new versions differed from the original ones for two items (see **Tables 3, 4**). Item 35 was present in the previous version of the N scale but it has not been included in the new one because of its redundant content. Reliability of the new scales largely resembles that of the original versions ( $\alpha = 0.83, 0.82$ ; TIF = 20.86, 20.80 for original and new N scale, respectively;  $\alpha = 0.73, 0.74$ ; TIF = 13.86, 14.15 for original and new L scale, respectively). Concerning N scale, a slight decrease of bias was observed (from 0.22 to 0.16). The other indexes remained substantially unchanged (bias = 0.20, 0.18 for original and new L scale, respectively; corrected correlation = 0.74, 0.75 for original and new L scale, respectively; 0.83, 0.84, for original and new N scale, respectively).

## Discussion

This study aimed at developing a new short version of the EPQ-R with improved psychometric characteristics. IRT based statistics allowed the identification of 48 items without gender DIF or misfit, well discriminating, and well distributed along the four latent traits continua. The new version of the P scale differs from the original one for eight items (out of 12), E scale for three, and N and L only for two. The largest improvement was reached for P scale, which in literature was found to perform less well than the

**TABLE 2** | Easiness ( $\epsilon$ ) and discrimination ( $\delta$ ) parameters for the 23 items of the Extraversion scale.

New form	Original form	Item n.	Text	$\epsilon$	$\delta$
	✓	47	Are you mostly quiet when you are with other people?	-5.623	-0.395
✓	✓	20	Do you enjoy meeting new people?	-1.914	1.613
	✓	11	Are you rather lively?	-1.605	1.311
✓	✓	58	Do you like mixing with people?	-1.592	2.182
		33	Do you prefer reading to meeting people?	-1.246	0.973
	✓	94	Do other people think of you as being very lively	-1.180	0.805
✓		28	Do you like going out a lot?	-1.105	1.193
		72	Do you often take on more activities than you have time for?	-0.832	0.570
		69	Do you often make decisions on the spur of the moment?	-0.755	0.489
		36	Do you have many friends?	-0.652	1.525
✓	✓	78	Can you get a party going?	-0.636	1.418
✓		67	Do you like doing things in which you have to act quickly?	-0.608	0.783
✓	✓	16	Can you usually let yourself go and enjoy yourself at a lively party?	-0.536	1.676
✓	✓	6	Are you a talkative person?	-0.468	1.211
		55	Do you like telling jokes and funny stories to your friends?	-0.365	0.705
		63	Do you nearly always have a 'ready answer' when people talk to you?	-0.230	0.401
✓	✓	45	Do you usually take the initiative in making new friends?	-0.171	1.741
✓	✓	24	Do you tend to keep in the background on social occasions?	0.151	1.658
✓	✓	51	Can you easily get some life into a rather dull party?	0.402	1.729
		40	Would you call yourself happy-go-lucky?	0.753	0.944
✓		61	Have people said that you sometimes act too rashly?	0.934	0.582
		1	Do you have many different hobbies?	0.974	0.519
✓	✓	90	Do you like plenty of bustle and excitement around you?	1.648	0.884

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

**TABLE 3** | Easiness ( $\epsilon$ ) and discrimination ( $\delta$ ) parameters for the 24 items of the Neuroticism scale.

New form	Original form	Item n.	Text	$\epsilon$	$\delta$
		87	Are you easily hurt when people find fault with you of the work you do?	-1.652	0.930
		97	Are you touchy about some things?	-1.528	0.696
✓	✓	22	Are your feelings easily hurt?	-1.486	1.143
		92	Are you sometimes bubbling over with energy and sometimes very sluggish?	-1.186	1.025
✓	✓	80	Do you worry too long after an embarrassing experience?	-1.139	1.128
		13	Do you often worry about things you should not have done or said?	-1.027	1.084
		43	Do you worry about awful things that might happen?	-0.819	0.497
		74	Do you worry a lot about your looks?	-0.460	0.484
✓	✓	3	Does your mood often go up and down?	-0.319	1.785
		100	When your temper rises, do you find it difficult to control?	-0.164	0.848
✓	✓	31	Are you often troubled about feelings of guilt?	-0.114	1.445
✓		65	Have you often felt listless and tired for no reason?	-0.003	1.720
	✓	8	Do you ever feel "just miserable" for no reason?	0.007	1.392
✓	✓	17	Are you an irritable person?	0.225	1.527
	✓	35	Would you call yourself a nervous person?	0.250	2.582
✓	✓	46	Would you call yourself tense or "highly-strung"?	0.308	2.928
✓	✓	84	Do you often feel lonely?	0.308	1.669
✓	✓	83	Do you suffer from "nerves"?	0.394	2.833
✓	✓	26	Do you often feel "fed-up"?	0.591	1.713
		76	Have you ever wished that you were dead?	0.976	1.102
✓		70	Do you often feel life is very dull?	0.990	1.273
		60	Do you suffer from sleeplessness?	2.443	0.691
✓	✓	38	Are you a worrier?	2.948	0.721
		52	Do you worry about your health?	8.462	-0.132

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

other three scales (e.g., Bishop, 1977; Block, 1977; Claridge, 1981). In particular, the new version is not affected by gender DIF and outperforms the original one for reliability and approximation of the scores obtained with the full-length form. The new versions of the other three scales performed as well as, or slightly better than the original ones. Although small in size, these improvements are valuable taking into account that were obtained by substituting a small number of items and reducing content redundancy.

## STUDY 2

This study aimed at investigating the functioning of the new version of the short EPQ-R on a new data set. Other to reliability and factor structure, construct validity was evaluated by taking into account relationships with social desirability, the dimensions of the FFM, and measures of anxiety and depression.

## Participants

Participants were 300 native Italian speakers aged between 18 and 65 (mean age = 29.28,  $SD = 10.38$ ; 60.2% females). They were recruited from different Italian regions using convenience sampling. All participants were presented with the new version of the short EPQ-R, whereas a subsample of 158 participants (mean

age = 34.73,  $SD = 9.88$ ; 68.7% females) also received the other measures. The participation to the study was anonymous and voluntary, and all standards for research with human subjects were respected. Written informed consent was obtained from the participants. The project has been approved, now as later, by the Ethical Committee for the Psychological Research of the University of Padova since a prospective ethics approval was not required at the time when the research was conducted (Protocol n. 2622).

## Instruments

The new form of the short EPQ-R devised in Study 1 was administered to all participants.

The five traits of the FFM of personality (i.e., extraversion, agreeableness, conscientiousness, emotional stability, and openness) were measured through the Italian version (Ubbiali et al., 2013; Chiorri et al., 2016) of the Big Five Inventory (BFI; John et al., 2008). The questionnaire consists of 44 items answered on a five-point Likert scale (from 1 "Strongly disagree" to 5 "Strongly agree"; e.g., "I see myself as someone who is full of energy" for extraversion; "I see myself as someone who is helpful and unselfish with others" for agreeableness; "I see myself as someone who perseveres until the task is finished"

**TABLE 4** | Easiness ( $\epsilon$ ) and discrimination ( $\delta$ ) parameters for the 21 items of the Lie scale.

New form	Original form	Item n.	Text	$\epsilon$	$\delta$
		98	Are you always willing to admit it when you have made a mistake?	-4.705	0.427
	✓	57	As a child were you ever cheeky to your parents?	-2.153	0.536
		62	Do you always wash before a meal?	-2.120	0.614
✓		4	Have you ever taken the praise for something you knew someone else had really done?	-2.057	0.678
✓	✓	19	Have you ever blamed someone for doing something you knew was really your fault?	-1.687	0.869
✓	✓	71	Have you ever taken advantage of someone?	-0.909	1.638
✓		32	Do you sometimes talk about things you know nothing about?	-0.733	1.188
	✓	15	If you say you will do something, do you always keep your promise no matter how inconvenient it might be?	-0.693	0.898
✓	✓	66	Have you ever cheated at a game?	-0.508	1.183
✓	✓	44	Have you ever broken or lost something belonging to someone else?	-0.172	1.299
✓	✓	27	Have you ever taken anything (even a pin or button) that belonged to someone else?	-0.149	1.382
		49	Do you sometimes boast a little?	0.164	0.996
✓	✓	10	Were you ever greedy by helping yourself to more than your share of anything?	0.303	1.201
✓	✓	53	Have you ever said anything bad or nasty about anyone?	0.419	1.513
		77	Would you dodge paying taxes if you were sure you could never be found out?	0.421	0.473
✓	✓	86	Do you always practice what you preach?	0.423	1.052
		89	Have you ever been late for an appointment or work?	0.451	1.094
		39	As a child did you do as you were told immediately and without grumbling?	0.691	0.717
✓	✓	23	Are all your habits good and desirable ones?	0.712	1.080
✓	✓	93	Do you sometimes put off until tomorrow what you ought to do today?	1.155	1.212
		82	Have you ever insisted on having your own way?	5.265	0.173

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

for conscientiousness; "I see myself as someone who worries a lot" for emotional stability; "I see myself as someone who is ingenious, a deep thinker" for openness). Convincing evidence was found concerning construct validity, factor structure, gender invariance, and reliability ( $\alpha$  from 0.75 to 0.86; Ubbiali et al., 2013; Chiorri et al., 2016;  $\alpha$  from 0.73 to 0.83 in the current sample).

The Impression Management (IM) scale of the Italian brief version (Bobbio and Manganelli, 2011) of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991) was also administered. The scale comprises 8 items answered on a six-point Likert scale (from 1 "Strongly disagree" to 6 "Strongly agree") and assesses the conscious tendency of individuals to provide positively inflated self-descriptions (e.g., "I have never dropped litter on the street"). Internal consistency of the scale ranges from 0.73 to 0.81 (Bobbio and Manganelli, 2011; in the current sample,  $\alpha = 0.75$ ).

The trait scale of the State-Trait Anxiety Inventory (STAI-Y; Spielberger et al., 1983; Pedrabissi and Santinello, 1989)

was used to evaluate anxiety. The scale comprises 20 items answered on a four-point Likert scale (from 1 "Not at all" to 4 "Very much"). The instrument evaluates the tendency of people to experience general anxiety and the relatively stable predisposition to view stressful situations as threatening (e.g., "I am regretful"). The Italian version of the questionnaire showed adequate validity and reliability ( $\alpha$  from 0.85 and 0.90; Pedrabissi and Santinello, 1989; in the current sample,  $\alpha = 0.92$ ).

Finally, the Italian version of the Patient Health Questionnaire-9 (PHQ-9; Spitzer et al., 1999; Kroenke et al., 2001) was used to evaluate depressive symptoms. The questionnaire is a self-administered instrument and assesses 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 through nine items scored on a four-point Likert scale (from 0 "Not at all" to 3 "Nearly every day"; e.g., "Feeling tired or having little energy"). This instrument showed adequate reliability ( $\alpha$  from

0.86 to 0.89), and good sensitivity and specificity (see Kroenke et al., 2001). In the current sample,  $\alpha$  equals 0.81.

## Analysis Strategy

Reliability of the new version of the short EPQ-R was tested through Cronbach's  $\alpha$ . Construct validity was evaluated by computing convergent validity coefficients and by analyzing the factor structure of the instrument.

Convergent validity was evaluated considering correlations between the four PEN-L traits, the five dimensions of FFM, social desirability, and indexes of depression and trait anxiety. According with literature, L scores are expected to positively correlate with the IM scale of the BIDR (e.g., Gillings and Joseph, 1996), while PEN traits are expected to correlate with BFI scales, depression and trait anxiety. In particular, positive correlations are expected between E scores of the EPQ-R and the extraversion measure of the BFI, while negative correlations are expected between P scale and agreeableness and conscientiousness. Positive correlations are also expected between N scale of the EPQ-R and the neuroticism measure of the BFI (e.g., McCrae and Costa, 1985; Draycott and Kline, 1995; Saggino, 2000; Barbaranelli et al., 2003; Scholte and De Bruyn, 2004; Heaven et al., 2013). Neuroticism, in addition, is expected to positively correlate with indexes of anxiety and depression (STAI-Y; Spielberger et al., 1983; PHQ-9; Spitzer et al., 1999; Kroenke et al., 2001). In contrast, extraversion is expected to negatively correlate with these two clinical indexes.

An Exploratory Structural Equation Model (ESEM; Asparouhov and Muthén, 2009) was run to evaluate the factor structure. The ESEM framework represents an integration of confirmatory factor analysis (CFA), structural equation modeling (SEM), and exploratory factor analysis (EFA). ESEMs give access to all the common statistics of SEM/CFA but, at the same time, overcome the restrictions associated with the confirmatory approach. CFA fixes non-target loadings to zero and, therefore, it may be inadequate to handle complex and multifaceted constructs where many cross-loadings may be expected (Marsh et al., 2009, 2010, 2011, 2014). When this is the case, fit problems and upward-biased estimates of correlations between factors can be observed (Cole et al., 2007; Marsh and Hau, 2007; Marsh et al., 2010). As in EFA, ESEMs allow for the free estimation of cross-loadings between items and non-target

factors. In this work, ESEM was run using Mplus7 (Muthén and Muthén, 2012), and the WLSMV as estimator (weighted least squares mean and variance-adjusted). This method is recommended for binary or ordinal observed data (e.g., Flora and Curran, 2004; Brown, 2006) such as the dichotomous items of the EPQ-R. In the model, the 48 items were the indicators and four factors were modeled. The GEOMIN oblique rotation was used. To evaluate the goodness of fit of the model, several fit indexes were considered:  $\chi^2$ , Comparative Fit Index (CFI; Bentler, 1990), Weighted Root Mean Square Residual (WRMR; Yu, 2002), and Root Mean Square Error of Approximation (RMSEA; Browne and Cudeck, 1993) with its 90% confidence interval (90% CI) and the test of close fit (CFit; Browne and Cudeck, 1993). A solution fits the data well when  $\chi^2$  is non-significant ( $p \geq 0.05$ ). Since this statistic is sensitive to sample size, the other fit measures were also considered. In particular, a solution fits the data well when CFI is close to 0.95 (0.90 to 0.95 for reasonable fit), WRMR is close to 1.0, and RMSEA is smaller than 0.06 (0.06 to 0.08 for reasonable fit) with CFit non-significant (see Hu and Bentler, 1999; Marsh et al., 2004; Brown, 2006).

## Results

Cronbach's  $\alpha$  coefficients were 0.55, 0.80, 0.81, and 0.70 for P, E, N, and L scales, respectively. These values were consistent with those of Study 1. Compared with the original version, the largest improvement was reached for P scale, as observed in Study 1.

Convergent validity coefficients are reported in **Table 5**. All the four PEN-L traits correlated in the expected direction with the considered constructs. E scale showed a strong positive relation with the extraversion measure of the BFI (0.727). P scale was negatively related to agreeableness ( $-0.323$ ) and conscientiousness ( $-0.321$ ). N scale was strongly correlated with neuroticism (0.709). Relations with anxiety and depression were also in the expected directions. N scale showed positive relations with scores of PHQ-9 (0.619) and STAI-Y (0.697), while moderate negative relations were found between these two indexes and E scale ( $r = -0.409, -0.405$  for PHQ-9 and STAI-Y, respectively). Finally, L scale showed a strong positive correlation with the IM scale of the BIDR.

Results of the ESEM supported the four-factor structure of the instrument ( $\chi^2_{(942)} = 1122.686, p < 0.001$ ; RMSEA = 0.025

**TABLE 5 |** Cronbach's  $\alpha$ s and correlations between the four PEN-L traits, STAI-Y, PHQ-9, BIDR-IM, and the five BFI dimensions.

	$\alpha$	Psychoticism	Extraversion	Neuroticism	Lie
STAI-Y	0.92	0.092	$-0.409^{***}$	0.697 <sup>***</sup>	$-0.259^{**}$
PHQ-9	0.81	0.099	$-0.405^{***}$	0.619 <sup>***</sup>	$-0.202^*$
BIDR-IM	0.75	$-0.429^{***}$	0.195*	$-0.293^{***}$	0.561 <sup>***</sup>
BFI-Extraversion	0.79	0.058	0.727 <sup>***</sup>	$-0.369^{***}$	0.132
BFI-Agreeableness	0.73	$-0.323^{***}$	0.317 <sup>***</sup>	$-0.241^{**}$	0.242 <sup>**</sup>
BFI-Conscientiousness	0.83	$-0.321^{***}$	0.220 <sup>**</sup>	$-0.385^{***}$	0.334 <sup>***</sup>
BFI-Neuroticism	0.80	0.021	$-0.363^{***}$	0.709 <sup>***</sup>	$-0.110$
BFI-Openness	0.78	0.094	0.262 <sup>**</sup>	$-0.126$	0.065

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



**TABLE 6** | Exploratory structural equation modeling.

Item	Psychoticism	Extraversion	Neuroticism	Lie
29	<b>0.279**</b>	-0.058	0.303**	-0.089
88	<b>0.317***</b>	0.066	0.101	-0.123
96	<b>0.444***</b>	0.065	0.017	0.050
41	<b>0.913***</b>	-0.076	-0.058	0.032
25	<b>0.384**</b>	-0.193	-0.119	-0.551***
42	<b>0.312**</b>	0.045	0.049	-0.304
37	<b>0.596***</b>	0.084	0.141	-0.255
64	<b>0.457***</b>	0.118	0.051	0.117
21	<b>0.728***</b>	-0.148	-0.050	0.058
30	<b>0.743***</b>	-0.236	-0.352*	-0.001
81	<b>0.546***</b>	0.223	0.286**	0.004
18	<b>0.531***</b>	-0.075	-0.063	-0.342**
16	-0.065	<b>0.796***</b>	-0.152	-0.141
78	0.083	<b>0.482***</b>	0.037	0.151
20	-0.081	<b>0.775***</b>	-0.067	0.176
51	0.182	<b>0.605***</b>	-0.014	0.014
67	0.075	<b>0.317***</b>	-0.052	-0.027
6	0.090	<b>0.374***</b>	-0.060	0.030
28	0.180	<b>0.678***</b>	0.033	0.154
58	-0.077	<b>0.865***</b>	-0.201	0.060
90	-0.004	<b>0.732***</b>	-0.017	-0.242
45	0.100	<b>0.600***</b>	0.020	0.100
24	-0.067	<b>0.817***</b>	-0.164	-0.094
61	0.396***	<b>0.499***</b>	0.414***	-0.038
65	-0.034	-0.162	<b>0.540***</b>	-0.184*
17	0.065	0.136	<b>0.701***</b>	0.187*
46	0.024	0.019	<b>0.844***</b>	0.245**
3	0.045	0.176	<b>0.798***</b>	-0.007
38	-0.217*	-0.151	<b>0.658***</b>	-0.016
80	-0.262*	-0.296*	<b>0.313***</b>	-0.040
26	-0.201*	-0.131	<b>0.683***</b>	-0.107
70	0.002	-0.191	<b>0.619***</b>	-0.070
31	-0.135	-0.140	<b>0.605***</b>	-0.029
22	-0.168	-0.114	<b>0.479***</b>	0.199*
38	-0.156	-0.210	<b>0.558***</b>	0.048
83	0.045	-0.034	<b>0.910***</b>	0.323
66	-0.094	-0.089	0.011	<b>0.664***</b>
27	-0.385*	0.047	-0.109	<b>0.318**</b>
32	0.018	-0.319**	-0.374***	<b>0.235*</b>
86	0.110	-0.057	-0.243**	<b>0.365***</b>
4	-0.402**	-0.005	-0.049	<b>0.483***</b>
93	0.000	-0.301**	-0.201*	<b>0.449***</b>
23	-0.144	0.069	-0.167	<b>0.369***</b>
53	0.060	0.043	-0.107	<b>0.639***</b>
10	0.054	0.050	-0.011	<b>0.566***</b>
44	-0.290*	-0.113	-0.066	<b>0.502***</b>
71	-0.359***	0.060	0.026	<b>0.476***</b>
19	-0.302**	-0.054	-0.326**	<b>0.191*</b>
CORRELATION WITH				
Extraversion	0.043			
Neuroticism	0.04	-0.198***		
Lie	-0.125	0.081	-0.258**	

Standardized factor loadings and factor correlations (N = 300).

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Bolded coefficients are target loadings.

[0.019, 0.031]; CFit  $\cong$  1.000; CFI = 0.930; WRMR = 0.864]. The model is represented in **Table 6**. All items loaded on the intended factor and cross-loadings were, in general, lower than those observed on the target-factor.

## Discussion

The analyses performed in this study provide further evidence concerning the adequate psychometric properties of the new short form of the EPQ-R. Concerning reliability, results are in line with those of Study 1 and confirm that, compared with the original version, the largest improvement was observed for P scale. Concerning validity, both the factor structure of the instrument and its convergent validity are supported.

## FINAL REMARKS

This work aimed at developing a new and improved version of the short form of the EPQ-R. This instrument is well-known and widely used in different settings. However, some weaknesses have been pointed out, especially for P scale (e.g., Bishop, 1977; Block, 1977; Claridge, 1981). IRT approach was used to develop the new instrument. This approach allowed for removing items with misfit or gender DIF, and for identifying items that were best at discriminating different levels of traits, while ensuring that the respective continua were covered. As suggested in literature, following these criteria for item selection should lead to a short scale with the same psychometric properties of the full-length instrument (Reise and Henson, 2000; Spence et al., 2012). In fact, results of this work show that the new short form of the EPQ-R approximated the scores obtained with the full-length form better than the original short version. In addition, convergent validity of the new scale was consistent with literature (e.g., Saklofske et al., 1995; Gillings and Joseph, 1996; del Barrio et al., 1997; Dazzi et al., 2004; Jylhä and Isometsä, 2006; Mor, 2010). The moderate to strong relationships between Eysenck's traits and clinical constructs provide further evidence toward the usefulness of assessing these traits in clinical settings.

A strength of the present work is that it provides a solution to some well-known drawbacks of the full-length EPQ-R and of its short form existing in the literature (Eysenck et al., 1985; Eysenck and Eysenck, 1991). The largest improvement was obtained for P scale. The new version is not affected by gender DIF and outperforms the original one for reliability and approximation of the full-length form. The new versions of the other three scales performed as well as the original ones, or slightly better. These improvements are small in size, yet notable considering that were obtained by substituting a small number of items and reducing content redundancy.

In the present work, separate analyses have been performed on each of the four scales by using a unidimensional IRT model. An alternative could have been examining the four scales at once through a multidimensional IRT (MIRT) model (see Haberman et al., 2008; Reckase, 2009). MIRT models offer some advantages over unidimensional IRT models. They could allow for better understanding the traits measured by an

instrument and how well individual items measure each of them (Ackerman, 1994). Moreover, MIRT models could provide a more precise estimation of scale reliability (Cheng et al., 2009) and item parameters (Finch, 2010). In the present work, some of these advantages are not very relevant. On the one hand, the factor structure of the EPQ-R has been widely tested and validated in the literature (e.g., Hosokawa and Ohyama, 1993; Maltby and Talley, 1998; Forrest et al., 2000; Qian et al., 2000; Scholte and De Bruyn, 2001; Aluja et al., 2003; Alexopoulos and Kalaitzidis, 2004; Dazzi et al., 2004; Francis et al., 2006; Karanci et al., 2006; Tiwari et al., 2009; Picconi et al., 2018). On the other hand, for scales whose length is analogous to that of the four EPQ-R scales (i.e., from 21 to 32 items), the unidimensional IRT models have been found to provide item parameter estimates whose precision exceeds or equals that of the estimates produced by the MIRT models (Finch, 2010). Finch (2010) investigated the precision of MIRT estimates on tests measuring a number of traits as small as two. For larger numbers of traits (e.g., the four traits of the EPQ-R), the number of parameters of a MIRT model increases considerably. Thus, the sample size of Study 1 (590 individuals) could have not been appropriate for performing a multidimensional analysis.

Concerning P scale, despite notable improvements, reliability remains rather low. This result, however, was expected. P scale, in fact, maybe because of its complex and clinical nature, is the most

problematic and controversial of the instrument (e.g., Eysenck et al., 1985). Future research, therefore, should try to develop a new pool of items effective in capturing the multifaced aspects of this trait.

In the present work, a new short version of the EPQ-R has been devised, which consists of 12 items per each of the four scales. An abbreviated form exists also in literature (Francis et al., 1992) that consists of only 6 items per scale. This abbreviated form suffers of the same weaknesses that have been pointed out for the other Eysenck's questionnaires. Future research should try to devise a new version of the abbreviated form by using the IRT approach.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

DC contributed to the conception and design of the study, conducted the research, performed the statistical analyses, and wrote the first draft of the manuscript. DC and PA wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

## REFERENCES

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Appl. Meas. Educ.* 7, 255–278. doi: 10.1207/s15324818ame0704\_1
- Alexopoulos, D. S., and Kalaitzidis, I. (2004). Psychometric properties of Eysenck Personality Questionnaire-Revised (EPQ-R) short scale in Greece. *Pers. Individ. Differ.* 37, 1205–1220. doi: 10.1016/j.paid.2003.12.005
- Almiro, P. A., Moura, O., and Simões, M. R. (2016). Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire—Revised (EPQ-R). *Pers. Individ. Differ.* 88, 88–93. doi: 10.1016/j.paid.2015.08.050
- Aluja, A., Garcia, Ó., and Garcia, L. F. (2003). A psychometric analysis of the revised Eysenck Personality Questionnaire short scale. *Pers. Individ. Differ.* 35, 449–460. doi: 10.1016/S0191-8869(02)00206-4
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th Ed.). Washington, DC: American Psychiatric Association.
- Anselmi, P., Vidotto, G., Bettinardi, O., and Bertolotti, G. (2015). Measurement of change in health status with Rasch models. *Health Qual. Life Out.* 13:16. doi: 10.1186/s12955-014-0197-x
- Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Modeling* 16, 397–438. doi: 10.1080/10705510903008204
- Baker, F. B. (2001). *The Basics of Item Response Theory* (2nd Ed.). Maryland: ERIC Clearinghouse on Assessment and Evaluation.
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsych. Dis. Treat.* 10, 153–165. doi: 10.2147/NDT.S53425
- Barbaranelli, C., Caprara, G. V., Rabasca, A., and Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Pers. Individ. Differ.* 34, 645–664. doi: 10.1016/S0191-8869(02)00051-X
- Barnes, G. E., Malamuth, N. M., and Check, J. V. (1984). Personality and sexuality. *Pers. Individ. Differ.* 5, 159–172.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246.
- Bergomi, M., Modenese, A., Ferretti, E., Ferrari, A., Licitra, G., Vivoli, R., et al. (2017). Work-related stress and role of personality in a sample of Italian bus drivers. *Work* 57, 433–440. doi: 10.3233/WOR-172581
- Bishop, D. V. (1977). The P scale and psychosis. *J. Abnorm. Psychol.* 86, 127–134.
- Blaszczynski, A. P., Buhrich, N., and McConaghy, N. (1985). Pathological gamblers, heroin addicts and controls compared on the EPQ 'Addiction Scale'. *Br. J. Addict.* 80, 315–319.
- Block, J. (1977). P scale and psychosis: continued concerns. *J. Abnorm. Psychol.* 86, 431–434.
- Bobbio, A., and Manganelli, A. M. (2011). Measuring social desirability responding. A short version of Paulhus' BIDR 6. *TPM* 18, 117–135. doi: 10.4473/TPM.18.2.4
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37, 29–51. doi: 10.1007/BF02291411
- Bock, R. D. (1997). A brief history of item response theory. *Educ. Meas. Issues Pract.* 16, 21–33. doi: 10.1111/j.1745-3992.1997.tb00605.x
- Bogaert, A. F. (1993). Personality, delinquency, and sexuality: data from three Canadian samples. *Pers. Individ. Differ.* 15, 353–356.
- Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., and de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. *Qual. Quant.* 47, 2341–2360. doi: 10.1007/s11135-012-9684-5
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.
- Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, eds. K. A. Bollen and J. S. Long (Newbury Park, CA: SAGE), 136–162.
- Cheng, Y. Y., Wang, W. C., and Ho, Y. H. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: a statistical solution for the bandwidth–fidelity dilemma. *Educ. Psychol. Meas.* 69, 369–388. doi: 10.1177/0013164408323241

- Chico, E., and Ferrando, P. J. (1995). A psychometric evaluation of the revised P scale in delinquent and non-delinquent Spanish samples. *Pers. Individ. Differ.* 18, 331–337.
- Chico, E., Tous, J. M., Lorenzo-Seva, U., and Vigil-Colet, A. (2003). Spanish adaptation of Dickman's impulsivity inventory: its relationship to Eysenck's personality questionnaire. *Pers. Individ. Differ.* 35, 1883–1892. doi: 10.1016/S0191-8869(03)00037-0
- Chiorri, C., Marsh, H. W., Ubbiali, A., and Donati, D. (2016). Testing the factor structure and measurement invariance across gender of the Big Five inventory through exploratory structural equation modeling. *J. Pers. Assess.* 98, 88–99. doi: 10.1080/00223891.2015.1035381
- Claridge, G. S. (1981). "Psychoticism," in *Dimensions of Personality. Papers in Honour of H.J. Eysenck*, eds. H. J. Eysenck, and R. Lynn (Oxford, UK: Pergamon Press), 79–110.
- Cohen, J. (1988). *Statistical Power and Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cole, D. A., Ciesla, J. A., and Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychol. Methods* 12, 381–398. doi: 10.1037/1082-989X.12.4.381
- Colledani, D. (2018). Psychometric properties and gender invariance for the Dickman Impulsivity Inventory. *Test. Psychometr. Methodol. Appl. Psychol.* 25, 49–61. doi: 10.4473/TPM25.1.3
- Colledani, D., Robusto, E., and Anselmi, P. (2018). Development of a new abbreviated form of the Junior Eysenck Personality Questionnaire-Revised. *Pers. Individ. Differ.* 120, 159–165. doi: 10.1016/j.paid.2017.08.037
- Conrad, P. J., Petersen, J. B., and Pihl, R. O. (1997). Disinhibited personality and sensitivity to alcohol reinforcement: independent correlates of drinking behavior in sons of alcoholics. *Alcohol. Clin. Exp. Res.* 21, 1320–1332.
- Corulla, W. J. (1990). A revised version of the psychoticism scale for children. *Pers. Individ. Differ.* 11, 65–76. doi: 10.1016/0191-8869(90)90169-R
- Da Dalt, L., Anselmi, P., Bressan, S., Carraro, S., Baraldi, E., Robusto, E., et al. (2013). A short questionnaire to assess pediatric resident's competencies: the validation process. *Ital. J. Pediatr.* 39:41. doi: 10.1186/1824-7288-39-41
- Da Dalt, L., Anselmi, P., Furlan, S., Carraro, S., Baraldi, E., Robusto, E., et al. (2015). Validating a set of tools designed to assess the perceived quality of training of pediatric residency programs. *Ital. J. Pediatr.* 41:2. doi: 10.1186/s13052-014-0106-2
- Dazzi, C. (2011). The Eysenck personality questionnaire-Revised (EPQ-R): A confirmation of the factorial structure in the Italian context. *Pers. Individ. Differ.* 50, 790–794. doi: 10.1016/j.paid.2010.12.032
- Dazzi, C., Pedrabissi, L., and Santinello, M. (2004). *Adattamento Italiano delle Scale di Personalità Eysenck per Adulti [Italian Adaptation of the Eysenck Personality Scales for Adults]*. Firenze, IT: Organizzazioni Speciali.
- del Barrio, V., Moreno-Rosset, C., López-Martínez, R., and Olmedo, M. (1997). Anxiety, depression and personality structure. *Pers. Individ. Differ.* 23, 327–335.
- Denney, D. R., and Frisich, M. B. (1981). The role of neuroticism in relation to life stress and illness. *J. Psychosom. Res.* 25, 303–307.
- Draycott, S. G., and Kline, P. (1995). The big three or the big five—The EPQ-R vs the NEOPI: a research note, replication and elaboration. *Pers. Individ. Differ.* 18, 801–804. doi: 10.1016/0191-8869(95)00010-4
- Escorial, S., and Navas, M. J. (2007). Analysis of the gender variable in the Eysenck Personality Questionnaire-revised scales using differential item functioning techniques. *Educ. Psychol. Meas.* 67, 990–1001. doi: 10.1177/0013164406299108
- Eysenck, H. J. (1991). Neuroticism, anxiety, and depression. *Psychol. Inq.* 2, 75–76.
- Eysenck, H. J., and Eysenck, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*. London, UK: University of London Press.
- Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Adult and Junior)*. London, UK: Hodder and Stoughton.
- Eysenck, H. J., and Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scale (Adults)*. London, UK: Hodder and Stoughton.
- Eysenck, S., and Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Pers. Individ. Differ.* 54, 485–489. doi: 10.1016/j.paid.2012.09.022
- Eysenck, S. B., Eysenck, H. J., and Barrett, P. (1985). A revised version of the psychoticism scale. *Pers. Individ. Differ.* 6, 21–29. doi: 10.1016/0191-8869(85)90026-1
- Eysenck, S. B. G., and Eysenck, H. J. (1963). On the dual nature of extraversion. *Brit. J. Clin. Psychol.* 2, 46–55.
- Ferrando, P. J. (2001). The measurement of neuroticism using MMQ, MPI, EPI and EPQ items: a psychometric analysis based on item response theory. *Pers. Individ. Differ.* 30, 641–656. doi: 10.1016/S0191-8869(00)00062-3
- Ferrando, P. J., and Anguiano-Carrasco, C. (2009). The interpretation of the EPQ Lie scale scores under honest and faking instructions: a multiple-group IRT-based analysis. *Pers. Individ. Differ.* 46, 552–556. doi: 10.1016/j.paid.2008.12.013
- Ferrando, P. J., and Chico, E. (2001). Detecting dissimulation in personality test scores: a comparison between person-fit indices and detection scales. *Educ. Psychol. Meas.* 61, 997–1012. doi: 10.1177/00131640121971617
- Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis-based models. *Appl. Psych. Meas.* 34, 10–26. doi: 10.1177/0146621609336112
- Flora, D. B., and Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods* 9, 466–491. doi: 10.1037/1082-989X.9.4.466
- Forrest, S., Lewis, C. A., and Shevlin, M. (2000). Examining the factor structure and differential functioning of the Eysenck Personality Questionnaire Revised-abbreviated. *Pers. Individ. Differ.* 29, 579–588. doi: 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. *Pers. Individ. Differ.* 21, 633–640. doi: 10.1016/0191-8869(96)00125-0
- Francis, L. J., Brown, L. B., and 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. *Pers. Individ. Differ.* 13, 443–449. doi: 10.1016/0191-8869(92)90073-X
- Francis, L. J., Lewis, C. A., and Ziebertz, H. G. (2006). The short-form revised Eysenck personality questionnaire (EPQR-S): a German edition. *Soc. Behav. Personal.* 34, 197–204. doi: 10.2224/sbp.2006.34.2.197
- Francis, L. J., and Pearson, P. R. (1988). Religiosity and the short-scale EPQ-R indices of E, N and L, compared with the JEPI, JEPQ and EPQ. *Pers. Individ. Differ.* 9, 653–651. doi: 10.1016/0191-8869(88)90162-6
- Gale, C. R., Booth, T., Möttus, R., Kuh, D., and Deary, I. J. (2013). Neuroticism and Extraversion in youth predict mental wellbeing and life satisfaction 40 years later. *J. Res. Pers.* 47, 687–697. doi: 10.1016/j.jrp.2013.06.005
- Gillings, V., and Joseph, S. (1996). Religiosity and social desirability: Impression management and self-deceptive positivity. *Pers. Individ. Differ.* 21, 1047–1050.
- Gómez-Benito, J., Hidalgo, M. D., and Padilla, J. L. (2009). Efficacy of effect size measures in logistic regression: an application for detecting DIF. *Methodology* 5, 18–25. doi: 10.1027/1614-2241.5.1.18
- Grau, E., and Ortet, G. (1999). Personality traits and alcohol consumption in a sample of non-alcoholic women. *Pers. Individ. Differ.* 27, 1057–1066.
- Gudgeon, E. T., Connor, J. P., Young, R. M., and Saunders, J. B. (2005). The relationship between personality and drinking restraint in an alcohol dependent sample. *Pers. Individ. Differ.* 39, 885–893. doi: 10.1016/j.paid.2005.03.009
- Haberman, S. J., von Davier, M., and 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., and Barkus, E. (2013). Agreeableness, conscientiousness, and psychoticism: distinctive influences of three personality dimensions in adolescence. *Brit. J. Psychol.* 104, 481–494. doi: 10.1111/bjop.12002
- Heaven, P. C., Newbury, K., and Wilson, V. (2004). The Eysenck psychoticism dimension and delinquent behaviours among non-criminals: changes across the lifespan? *Pers. Individ. Differ.* 36, 1817–1825. doi: 10.1016/j.paid.2003.07.003
- Hosokawa, T., and Ohyama, M. (1993). Reliability and validity of a Japanese version of the short-form Eysenck Personality Questionnaire—Revised. *Psychol. Rep.* 72, 823–832.
- Howarth, E. (1986). What does Eysenck's psychoticism scale really measure? *Brit. J. Psychol.* 77, 223–227.
- Hoyle, R. H., Fejfar, M. C., and Miller, J. D. (2000). Personality and sexual risk taking: a quantitative review. *J. Pers.* 68, 1203–1231. doi: 10.1111/1467-6494.00132
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* 6, 1–55.



- Huang, L., Zhou, D., Yao, Y., and Lan, Y. (2015). Relationship of personality with job burnout and psychological stress risk in clinicians. *Chin. J. Industr. Hyg. Occup. Dis.* 33, 84–87.
- Jodoin, M. G., and Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Appl. Meas. Educ.* 14, 329–349. doi: 10.1207/S15324818AME1404\_2
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). “Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues,” in *Handbook of Personality: Theory and Research*, eds. O.P. John, R.W. Robins, and L.A. Pervin (New York, NY: Guilford), 114–158.
- Judge, T. A., Bono, J. E., and Locke, E. A. (2000). Personality and job satisfaction: the mediating role of job characteristics. *J. Appl. Psychol.* 85, 237–249. doi: 10.1037/0021-9010.85.2.237
- Jylhä, P., and Isometsä, E. (2006). The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depress. Anxiety* 23, 281–289. doi: 10.1002/da.20167
- Karancı, A. N., Dirik, G., and Yorulmaz, O. (2006). [Reliability and validity studies of Turkish translation of Eysenck personality questionnaire revised-Abbreviated]. *Turk. J. Psychiatry* 18, 254–261.
- Katz, Y. J., and Francis, L. J. (2000). Hebrew revised Eysenck Personality Questionnaire: Short form (EPQR-S) and abbreviated form (EPQR-A). *Soc. Behav. Personal.* 28, 555–560. doi: 10.2224/sbp.2000.28.6.555
- Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x
- Laidra, K., Pullmann, H., and Allik, J. (2007). Personality and intelligence as predictors of academic achievement: a cross-sectional study from elementary to secondary school. *Pers. Individ. Differ.* 42, 441–451. doi: 10.1016/j.paid.2006.08.001
- Lajunen, T., and Scherler, H. R. (1999). Is the EPQ Lie Scale bidimensional? Validation study of the structure of the EPQ Lie Scale among Finnish and Turkish university students. *Pers. Individ. Differ.* 26, 657–664.
- Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *J. Clin. Psychol.* 23, 84–86. doi: 10.1002/1097-4679(196701)23:1<84::AID-JCLP2270230123>3.0.CO;2-2
- Lodhi, P. H., and Thakur, S. (1993). Personality of drug addicts: Eysenckian analysis. *Pers. Individ. Differ.* 15, 121–128.
- Lu, L. (1995). The relationship between subjective well-being and psychosocial variables in Taiwan. *J. Soc. Psychol.* 135, 351–357.
- Lynn, R., and Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *J. Soc. Psychol.* 137, 369–373.
- Magis, D., Beland, S., and Raiche, G. (2016). Package ‘difR’. R package version 4.7.
- Maij-de Meij, A. M., Kelderman, H., and Flier, V. D. H. (2008). Fitting a mixture item response theory model to personality questionnaire data: characterizing latent classes and investigating possibilities for improving prediction. *Appl. Psych. Meas.* 32, 611–631. doi: 10.1177/0146621607312613
- Maltby, J., and Talley, M. (1998). The psychometric properties of an abbreviated form of the revised Junior Eysenck Personality Questionnaire (JEPQR-A) among 12–15-year old US young persons. *Pers. Individ. Differ.* 24, 891–893. doi: 10.1016/S0191-8869(97)00234-1
- Marsh, H. W., and Hau, K. T. (2007). Applications of latent-variable models in educational psychology: the need for methodological-substantive synergies. *Contemp. Educ. Psychol.* 32, 151–170. doi: 10.1016/j.cedpsych.2006.10.008
- Marsh, H. W., Hau, K. T., and 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. *Struct. Equ. Modeling* 11, 320–341. doi: 10.1207/s15328007sem1103\_2
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., et al. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychol. Assess.* 22, 471–491. doi: 10.1037/a0019227
- Marsh, H. W., Morin, A. J., Parker, P. D., and Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psycho.* 10, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students’ evaluations of university teaching. *Struct. Equ. Modeling* 16, 439–476. doi: 10.1080/10705510903008220
- Marsh, H. W., Nagengast, B., Morin, A. J., Parada, R. H., Craven, R. G., and Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: an application of exploratory structural equation modeling. *J. Educ. Psychol.* 103, 701–732. doi: 10.1037/a0024122
- McCrae, R. R., and Costa, P. T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Pers. Individ. Differ.* 6, 587–597. doi: 10.1016/0191-8869(85)90008-X
- Mor, N. (2010). “Eysenck personality questionnaire,” in *The Corsini Encyclopedia of Psychology*, eds. I. B. Weiner and W. E. Craighead (Hoboken, NJ: John Wiley & Sons, Inc.), 1–2.
- Muthén, L. K., and Muthén, B. O. (2012). *Mplus Version 7 User’s Guide*. Los Angeles, CA: Muthén & Muthén.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692. doi: 10.1093/biomet/78.3.691
- Nyborg, H. E. (1997). *The Scientific Study of Human Nature: tribute to Hans J. Eysenck at Eighty*. New York, NY: Elsevier.
- Paulhus, D. L. (1991). “Measurement and control of response bias,” in *Measures of Personality and Social Psychological Attitudes*, Vol. 1, eds. J. P. Robinson, P. R. Shaver, and L. S. Wrightsman (San Diego, CA: Academic Press), 17–59.
- Pedrabissi, L., and Santinello, M. (1989). *Inventario per Lansia 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., and 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 Health* 18, 25–34. doi: 10.1016/j.jval.2014.10.005
- Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., and Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck Personality Questionnaire-Short form (EPP-S). *Pers. Individ. Differ.* 123, 76–80. doi: 10.1016/j.paid.2017.11.013
- Qian, M., Wu, G., Zhu, R., and Zhang, S. (2000). Development of the revised Eysenck personality questionnaire short scale for Chinese (EPQ-RSC). *Acta Psychol. Sinica* 32, 317–323.
- R Core Team (2016). *A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Available online at: <http://www.R-project.org/>
- Reckase, M. D. (2009). *Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences*. New York, NY: Springer. doi: 10.1007/978-0-387-89976-3
- Reise, S. P. (1990). A comparison of item-and person-fit methods of assessing model-data fit in IRT. *Appl. Psych. Meas.* 14, 127–137. doi: 10.1177/014662169001400202
- Reise, S. P., and Henson, J. M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment* 7, 347–364. doi: 10.1177/107319110000700404
- Rizopoulos, D. (2012). Package ‘ltm’. R package version 1.1–0.
- Rocklin, T., and Revelle, W. (1981). The measurement of extroversion: a comparison of the Eysenck Personality Inventory and the Eysenck Personality Questionnaire. *Brit. J. Soc. Psychol.* 20, 279–284.
- Roger, D., and Morris, J. (1991). The internal structure of the EPQ scales. *Pers. Individ. Differ.* 12, 759–764.
- Saggino, A. (2000). The big three or the big five? A replication study. *Pers. Individ. Differ.* 28, 879–886. doi: 10.1016/S0191-8869(99)00146-4
- Saklofske, D. H., Kelly, I. W., and Janzen, B. L. (1995). Neuroticism, depression, and depression proneness. *Pers. Individ. Differ.* 18, 27–31.
- Sanavio, E., Bertolotti, G., Bettinardi, O., Michielin, P., Vidotto, G., and Zotti, A. M. (2013). The cognitive behavioral assessment (CBA) project: presentation and proposal for international collaboration. *Psychol. Community Health* 2, 362–380. doi: 10.5964/pch.v2i3.61
- Scholte, R. H., and De Bruyn, E. E. (2001). The Revised Junior Eysenck Personality Questionnaire (JEPQ-R): Dutch replications of the full-length, short, and abbreviated forms. *Pers. Individ. Differ.* 31, 615–625. doi: 10.1016/S0191-8869(00)00166-5
- Scholte, R. H., and De Bruyn, E. E. (2004). Comparison of the Giant Three and the Big Five in early adolescents. *Pers. Individ. Differ.* 36, 1353–1371. doi: 10.1016/S0191-8869(03)00234-4

- Smillie, L. D., Bhairo, Y., Gray, J., Gunasinghe, C., Elkin, A., McGuffin, P. et al. (2009). Personality and the bipolar spectrum: normative and classification data for the Eysenck Personality Questionnaire-Revised. *Compr. Psychiat.* 50, 48–53. doi: 10.1016/j.comppsy.2008.05.010
- Sotgiu, I., Anselmi, P., and Meneghini, A. M., (2018). Investigating the psychometric properties of the Questionnaire for Eudaimonic Well-Being: A Rasch analysis. *Test. Psychometr. Methodol. Appl. Psychol.* (in press).
- Spence, R., Owens, M., and Goodyer, I. (2012). Item response theory and validity of the NEO-FII in adolescents. *Pers. Individ. Differ.* 53, 801–807. doi: 10.1016/j.paid.2012.06.002
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., and Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory STAI*. Palo Alto, CA: Consulting Psychologists Press.
- Spitzer, R. L., Kroenke, K., and Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 282, 1737–1744.
- Thissen, D., and Steinberg, L. (2009). “Item response theory,” in *The Sage Handbook of Quantitative Methods in Psychology*, eds. R. Millsap and A. Maydeu-Olivares (London: Sage), 148–177.
- Tiwari, T., Singh, A. L., and Singh, I. L. (2009). The short-form revised Eysenck personality questionnaire: a Hindi edition (EPQRS-H). *Ind. Psychiatry J.* 18, 27–31. doi: 10.4103/0972-6748.57854
- Ubbiali, A., Chiorri, C., and Hampton, P. (2013). Italian Big Five Inventory. Psychometric properties of the Italian adaptation of the Big Five Inventory (BFI). *Appl. Psychol. Bull.* 266, 37–48.
- Vidotto, G., Cioffi, R., Saggino, A., and Wilson, G. (2008). The Italian version of the Junior Eysenck Personality Questionnaire: a confirmatory factor analysis. *Psychol. Rep.* 103, 715–726. doi: 10.2466/pr0.103.3.715-726
- Wood, J., and Newton, A. K. (2003). The role of personality and blame attribution in prisoners' experiences of anger. *Pers. Individ. Differ.* 34, 1453–1465. doi: 10.1016/S0191-8869(02)00127-7
- Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Meas. Trans.* 8:370.
- Wright, B. D., and Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: Mesa Press.
- Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes (Doctoral dissertation, University of California Los Angeles)*. Available online at: [http://www.researchgate.net/profile/Brian\\_Goldman3/publication/222578792\\_A\\_Multicomponent\\_Conceptualization\\_of\\_Authenticity\\_Theory\\_and\\_Research/links/00b49520552576cb3b000000.pdf](http://www.researchgate.net/profile/Brian_Goldman3/publication/222578792_A_Multicomponent_Conceptualization_of_Authenticity_Theory_and_Research/links/00b49520552576cb3b000000.pdf)
- Zanon, C., Hutz, C. S., Yoo, H., and Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica* 29:18. doi: 10.1186/s41155-016-0040-x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Colledani, Anselmi and Robusto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.