TPM®

CROSS-CULTURAL VALIDATION OF A NEW ABBREVIATED VERSION OF THE EPQ-R

DAIANACOLLEDANI
PASQUALE ANSELMI
EGIDIO ROBUSTO
UNIVERSITY OF PADOVA

The present work aims at providing evidence concerning the psychometric properties of a new abbreviated version of the Eysenck Personality Questionnaire-Revised (EPQ-R) in a cross-cultural sample of native English speakers (recruited in three geographical areas: North America, Europe, and Oceania. The four-factor structure of the questionnaire was confirmed, as well as the satisfactory reliability and convergent validity of its scales. Moreover, item-level analyses showed that the items of the scales were simple structured, without misfit, and without cultural, age, and gender biases. On the whole, the results suggest the suitability of the new abbreviated version of the EPQ-R in English contexts.

Keywords: EPQR-A; Two-parameter logistic model (2PL); Differential item functioning; Big Five.

Correspondence concerning this article should be addressed to Daiana Colledani, FISPPA Department – Section of Applied Psychology, University of Padova, Via Venezia 14, 35131, Padova (PD), Italy. Email: daiana.colledani@unipd.it

The contribution of Eysenck in the study of personality is one of the most influential of the last 50 years (Boyle et al., 2008). Based on his extensive research, Eysenck came to devise a model of personality which includes three basic traits, Psychoticism (P), Extraversion (E), and Neuroticism (N), and is often referred to as the "Giant Three" or PEN model.

For nearly 40 years, Hans and Sybil Eysenck developed, updated, validated, and refined their personality measures (Boyle et al., 2008). The last revision of their instruments consists of 100 items and includes four scales, three for the assessment of PEN traits and a Lie (L) scale that detects social desirability bias (Eysenck et al., 1985). Short and abbreviated versions of the instrument, that assess the four dimensions through 48 and 24 items respectively (12 or 6 items for each scale), were also developed. These brief versions demonstrated acceptable psychometric properties and obtained great approval in the cross-cultural, scientific, and professional fields (Eysenck & Barrett, 2013; McLarnon & Romero, 2020), due to their usefulness in the assessment of personality when time is limited. However, some criticisms have been raised, mainly concerning the P scale (low range of scoring, skewed positive distribution, and low internal consistency maybe due to a multifaceted structure). Moreover, several studies indicated that some items of P, N, and L scales exhibit differential item functioning (DIF) across gender (e.g., Colledani, Anselmi, & Robusto, 2018; Escorial & Navas, 2007; Karanci et al., 2007).

Recently, a new abbreviated version of the Eysenck Personality Questionnaire-Revised (EPQ-R) has been developed in the Italian context, that aimed to reduce these shortcomings (Colledani et al., 2019a). The authors used item-level statistics and procedures introduced within the framework of multidimensional item response theory (MIRT), that have been proved to be useful for the development of abbreviated instruments (Anselmi et al., 2015; Bock, 1997; Colledani, 2018; Colledani, Robusto, & Anselmi, 2018; Colledani et al., 2019b; Haberman et al., 2008; Reckase, 2009; Thissen & Steinberg, 2009). These

methods allowed for selecting, from the full-length version of the instrument, 24 items with simple structure, good discrimination, and coverage of the latent trait continua, and without gender DIF and misfit. The scales developed with this procedure were found to outperform the previous ones in reliability (even if the P scale remained the most problematic) and approximation of the measures obtained with the full-length test.

The present work aims to investigate the functioning of the new abbreviated version of the EPQ-R developed in the Italian context on a cross-cultural sample of native English speakers recruited in three geographical areas (i.e., North America, Europe, and Oceania).

Method

Participants and Procedure

A total of 412 native English speakers (females = 223; mean age = 33.22, *SD* = 11.91, from 18 to 82 years) were recruited via Prolific Academic (http://prolific.ac) from Oceania (Australia and New Zealand; *N* = 106), North America (USA and Canada; *N* = 142), and Europe (UK and Ireland; *N* = 151; nationality was missing for 13 participants). Prolific Academic is a platform for data collection which permits recruiting large and diverse samples from all around the world. Thus, it is often used to carry out multicultural studies (Colledani & Camperio Ciani, 2021; Palan & Schitter, 2018; Sauter et al., 2020). On this platform, participants know that they are recruited to participate in research, are aware of the expected payments, treatment, rights, and obligations. Participation in the study was anonymous and voluntary. All participants filled out a set of self-report questionnaires available after agreeing with an electronic informed consent and received £ 0.90 in exchange for participation. The mean time taken to complete the task was 6.47 minutes, the dropout rate was low (3.16%, *N* = 13 dropouts), and the participants' mean Prolific Score was 99.95 (this is a score assigned by Prolific and based on the quality of the participants' performance in previously completed studies on the platform; the more accurate their submissions, the closer their prolific score is to 100).

Measures

The English version of the abbreviated form of the EPQ-R developed by Colledani et al. (2019a) consists of 24 items extracted from the full-length version of the instrument (Eysenck et al., 1985). Each of the four PEN-L scales includes six dichotomous items (yes/no). A short form of the Big Five Inventory (Rammstedt & John, 2007) was administered, which includes two items for each scale, scored on a 5-point Likert scale (from 1 *Strongly disagree* to 5 *Strongly agree*). The eight items of the Italian version of the impression management (IM) subscale of the Balanced Inventory of Desirable Responding (BIDR; Bobbio & Manganelli, 2011) were employed to assess social desirability. Answers were scored on a 6-point scale (from 1 *Strongly disagree* to 6 *Strongly agree*). In the current sample, Cronbach's α was .72.

The Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) and General Anxiety Disorder-7 (GAD-7; Plummer et al., 2016) were administered to evaluate depression and anxiety. These instruments include nine and seven items, respectively, asking participants to evaluate, on a 4-point frequency scale, the presence of depression and anxiety symptoms over the last two weeks. In the current sample, Cronbach's αs were .92 and .93 for PHQ-9 and GAD-7, respectively. Three items were used to assess satisfaction with social

relationships (e.g., "I feel satisfied about my social relationships). The items were scored on a 5-point Likert scale (from 1 = *strongly disagree* to 5 = *strongly agree*). In the current sample, Cronbach's α was .83.

Finally, eight items were used to evaluate the frequency of engagement in risky behaviors or use of substances in the last six months. Specifically, four of these items investigated the frequency with which participants smoked cigarettes, had occasional unprotected sex, law problems, and performed risky driving behaviors; whereas the remaining four items assessed the use of illegal drugs (i.e., "smoking joints," "using drugs"), and the use of psychopharmacologic drugs (i.e., "Anxiolytics/Benzodiazepines," "Psychopharmacological substances"). The responses to these four items were averaged to obtain two composite scores: one pertaining to the use of illegal drugs, and the other pertaining to the use of psychopharmacologic drugs. These eight items were scored on a 5-point scale (from 1 = *never* to 5 = *very often*).

All these instruments have been found to have satisfactory psychometric properties and have been used in recent studies (e.g., Colledani et al., 2019a; Colledani, Capozza et al., 2018; Kircaburun & Griffiths, 2018; McLarnon & Romero, 2020; Moreno et al., 2019).

Analysis Strategy

The analytic procedure that Colledani et al. (2019a) used for the development of this abbreviated form of the EPQ-R was replicated. Methods and statistics proposed within the framework of MIRT were used to evaluate DIF and misfit. To detect DIF across gender, age (i.e., 18-39 years, 40-82 years), and cultures (i.e., North America, Europe, and Oceania), multiple-group confirmatory 2PL MIRT models were used and the invariance of easiness (uniform bias) and discrimination (nonuniform bias) parameters was tested through the Wald test. To detect misfitting items, the signed chi-squared test (S-χ^2; Orlando & Thissen, 2000) was employed. Since S-χ^2 and Wald tests are sensitive to the sample size, effect size measures (Cohen, 1988) were computed to detect noticeable misfit and DIF ($\Phi \leq .30$ defines negligible size; $.30 \leq \Phi < .50$ defines medium effect size; and $\Phi \geq .50$ defines large size). These analyses were run with the R package "mirt" (Chalmers et al., 2018).

An exploratory structural equation model (ESEM) approach was used to test the factor structure of the scale. The model was run using Mplus7 (Muthén & Muthén, 2012) and the weighted least squares mean and variance-adjusted (WLSMV) estimator. The reliability of the abbreviated EPQ-R scales was evaluated through Cronbach's α and McDonald's (1999) ω, whereas convergent validity was investigated considering correlations of PEN-L scores with Big Five, impression management, psychosocial and behavioral measures.

Results

The ESEM showed an excellent fit: $\chi^2(186) = 215.66$, $p = .067$; CFI = .99; RMSEA = .020, [.000, .030], $p = 1.000$; SRMR = .058. All items significantly loaded on the intended factor (loadings from .32 to .88; see Table 1), and only two items showed meaningful cross-loadings (i.e., loadings ≥ .32 on more than one factor; Tabachnick et al., 2007; Item 11 of P scale and Item 6 of L scale had cross-loadings on L and E scales, respectively). Several items showed misfit (i.e., five items for P, E, and N scales, and two items for L scale). However, the effect size of misfit was negligible for all of them (see Table 1). Concerning gender and age DIF, no item showed uniform or nonuniform biases (Items 9, 2, and 18 for P, E, and L scales, re-

spectively, exhibited uniform gender bias of negligible size; Item 1 of N scale exhibited uniform age bias of negligible size). Also for cultural DIF no item exhibited uniform or nonuniform biases (Items 2, 4, 14, 16, and 17 of E scale, and Item 24 of P scale showed uniform bias of negligible size). Thus, results suggest that the four scales of the English version of the abbreviated EPQ-R by Colledani et al. (2019a) were simple structured and composed of well-fitting and unbiased items (for gender, age, and culture).

TABLE 1
Factor loadings of the ESEM model, DIF, and item fit statistics

					Gender				Age				Culture						
					Intercept		Slope		Intercept		Slope		Intercept		Slope				
Item	λP	λE	λN	λL	Wald	ES	Wald	ES	Wald	ES	Wald	ES	Wald	ES	Wald	ES	$S\chi^2$	*df*	ES
9	.61***	–.03	–.06	–.07	13.55**	.18	1.92	.07	0.17	.02	0.59	.04	1.83	.07	0.02	.01	20.87**	6	.23
11	.42***	.12	.03	**–.48*** **	2.61	.08	0.90	.05	0.37	.03	0.11	.02	0.05	.01	0.30	.03	13.50**	4	.18
12	.82***	–.22	.01	.13	0.37	.03	0.31	.03	0.07	.01	0.29	.03	0.40	.03	0.56	.04	14.78***	1	.19
13	.71***	.03	.10	–.22	1.85	.07	0.45	.03	0.17	.02	0.32	.03	0.35	.03	0.22	.02	21.32**	6	.23
21	.52***	.05	.04	–.11	0.50	.04	0.85	.05	0.24	.02	0.01	.00	0.56	.04	1.10	.05	9.31	5	.15
24	.32***	–.25**	.08	.08	0.19	.02	0.18	.02	3.70	.09	1.28	.06	7.24*	.13	0.08	.01	13.14*	6	.18
2	–.08	.69***	.10	.09	10.92**	.16	0.22	.02	4.70	.11	0.50	.04	11.00*	.17	0.57	.04	12.56	6	.17
4	.06	.68***	–.19*	–.06	0.61	.04	0.02	.01	0.09	.02	0.32	.03	7.77*	.14	0.75	.04	15.84*	6	.20
14	–.04	.77***	–.05	.00	0.01	.00	0.07	.01	4.46	.10	0.02	.01	6.32*	.13	0.41	.03	15.34**	5	.19
16	.20*	.80***	–.08	.21*	0.01	.01	0.58	.04	0.12	.02	0.47	.03	6.82*	.13	0.23	.02	22.51***	5	.23
23	.02	.75***	.09	–.12	1.00	.05	0.30	.03	7.26	.13	0.59	.04	5.64	.12	0.13	.02	13.01*	5	.18
17	–.15	.88***	.08	–.02	0.58	.04	0.73	.04	0.02	.01	1.12	.05	6.42*	.13	0.05	.01	18.69**	5	.21
1	.09	.08	.78***	.12	2.84	.08	1.00	.05	15.32***	.19	1.67	.06	3.47	.09	0.59	.04	18.16***	4	.21
5	.06	–.03	.60***	–.09	2.63	.08	0.09	.01	4.66	.11	0.01	.01	2.89	.09	1.30	.06	10.05	6	.16
7	.00	.04	.82***	–.02	0.51	.04	2.10	.07	3.73	.10	0.04	.01	1.09	.05	0.77	.04	13.35*	5	.18
10	.04	.11	.48***	–.07	3.95	.10	0.17	.02	0.46	.03	0.66	.04	0.63	.04	0.18	.02	19.60**	6	.22
15	–.10	–.09	.75***	.06	4.30	.10	3.12	.09	1.95	.07	0.21	.02	0.11	.02	0.74	.04	13.47*	5	.18
19	.11	–.18**	.70***	–.05	2.89	.08	0.04	.01	6.08	.12	0.85	.05	1.26	.06	0.10	.02	12.13*	5	.17
3	.12	.00	–.15	.71***	2.99	.09	0.87	.05	0.25	.02	0.18	.02	0.00	.00	0.02	.01	8.27	5	.14
6	.02	.26**	**–.39*** **	.40***	0.76	.04	0.03	.01	3.11	.09	0.20	.02	2.36	.08	0.35	.03	9.37	5	.15
8	–.01	–.01	.10	.65***	3.14	.09	0.26	.03	2.10	.07	2.24	.07	0.65	.04	1.25	.06	10.06	5	.16
18	–.17	.01	.14	.62***	11.74**	.17	3.85	.10	6.07	.12	1.62	.06	1.10	.05	0.20	.02	21.76***	6	.23
20	–.05	–.01	.05	.82***	1.42	.06	0.51	.04	0.69	.04	1.34	.06	0.20	.02	0.27	.03	20.85***	5	.22
22	.04	.16*	–.05	.36***	0.38	.03	1.02	.05	4.20	.10	1.19	.05	0.01	.00	0.93	.05	9.30	6	.15

Note. ESEM = exploratory structural equation model; DIF = differential item functioning; λP, λE, λN, λL = factor loading of the ESEM model for P (Psychoticism), E (Extraversion), N (Neuroticism), and L (Lie) scales, respectively; S-χ^2 = item fit index; *df* = degrees of freedom of S-χ^2; ES = effect size; Slope represents nonuniform bias; Intercept represents uniform bias. Correlations ranged from –.19 (between N and L), $p < .05$, to .24 (between P and N), $p < .05$. In bold are reported meaningful cross-loadings (i.e., loadings ≥ .32).
*$p < .05$. **$p < .01$. ***$p < .001$.

Cronbach's αs, for the total sample, were .50, .79, .72, and .61 whereas McDonald's ωs were .61, .92, .77, and .69 for P, E, N, and L scales, respectively (see Table 2 for results distinct by gender, age, and nationality). Cronbach's αs are in line with those found by Colledani et al. (2019a; αs =.54, .65, .72, and .61 for PEN-L, respectively) in the Italian context.

TABLE 2
Cronbach's α and McDonald's ω

		Psychoticism		Extraversion		Neuroticism		Lie	
	N	α	ω	α	ω	α	ω	α	ω
Total sample	412	.50	.61	.79	.92	.72	.77	.61	.69
Female	223	.48	.58	.77	.81	.71	.77	.58	.66
Male	182	.48	.60	.80	.84	.73	.76	.64	.71
Age class (18-39 years)	307	.49	.61	.79	.83	.70	.73	.58	.66
Age class (40-82 years)	103	.51	.57	.79	.84	.72	.78	.68	.71
Oceania	106	.47	.55	.79	.84	.75	.81	.60	.71
North America	142	.49	.60	.72	.74	.74	.77	.63	.71
Europe	151	.56	.65	.81	.84	.65	.71	.60	.66

Correlations between the PEN scales and the Big Five measures supported convergent validity (see Table 3). In fact, P scale showed a moderate negative correlation with agreeableness and conscientiousness (*r*s = –.22, –.26, *p*s < .001), whereas E and N were strongly and positively correlated with the Extraversion and Neuroticism measures of the Big Five, respectively (*r*s = .75, .55, *p*s < .001). A strong positive correlation was also found between the L scale and the IM subscale of the BIDR (*r* = .60, *p* < .001). Also correlations with psychosocial and behavioral measures supported convergent validity. In fact, P was positively correlated with the measures of law problems, use of substances, anxiety, depression, and risky behaviors (*r* ranging from .12 to .28, *p*s < .05, < .001), whereas it was negatively correlated with satisfaction with social relationships (*r* = –.30, *p* < .001). Neuroticism was strongly and positively associated with anxiety and depression (*r*s = .64 and .62, respectively; *p*s < .001), whereas Extraversion was positively and strongly associated with satisfaction with social relationships (*r* = .40, *p* < .001) and weakly associated with sensation seeking behaviors (illegal drug use, cigarette smoking, risky driving behaviors; *r*s ranging from .10 to .15, *p*s < .05, < .01).

TABLE 3
Correlations of PEN-L scores with BFI measures, social desirability, satisfaction with social relations, GAD-7, PHQ-9, and behavioral measures

	Psychoticism	Extraversion	Neuroticism	Lie
Agreeablesness-BFI	–.22***	.19***	–.42***	.17***
Coscientiousness-BFI	–.26***	.14**	–.23***	.27***
Extraversion-BFI	–.00	.75***	–.21***	.03
Neuroticism-BFI	.07	–.28***	.55***	–.15**
Openness-BFI	.02	.05	.08	–.05
IM-BIDR	–.40***	–.04	–.25***	.60***
Satisfaction with social relations	–.30***	.40***	–.34***	.07
GAD-7	.19***	–.13**	.64***	–.18***
PHQ-9	.28***	–.15**	.62***	–.16**
Psychopharmacological drugs	.12*	.03	.18***	–.10*

(Table 3 continues)

Table 3 (continued)

	Psychoticism	Extraversion	Neuroticism	Lie
Illegal drugs	.26***	.15**	.13**	–.09
Cigarettes	.18***	.10*	.10*	–.01
Unprotected sex	.12*	.09	.11*	–.04
Risky driving behaviors	.12*	.11*	.09	–.25***
Law problems	.19***	.09	.08	–.04

Note. PEN-L = Psychoticism, Extraversion, Neuroticism, and Lie; BFI = Big Five Inventory; IM-BIDR = Impression management-Balanced Inventory of Desirable Responding; GAD-7 = General Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This work aimed to test the psychometric properties of the new abbreviated version of the EPQ-R, developed in the Italian context (Colledani et al. 2019a), on a cross-cultural English speaking sample from Oceania, Europe, and North America.

The work showed that the instrument has adequate psychometric properties in the different cultural groups considered. The results of the analyses confirmed the four-factor structure of the instrument and showed that the 24 items are well-fitting, simple structured (excluding Items 6 and 11 of the L scale and P scale, respectively), and without gender, age, and cultural DIF. The correlations of the PEN-L scores with Big Five, BIDR, psychosocial and behavioral measures were in line with expectations, thus supporting convergent validity. Moreover, all scales showed satisfactory to good reliability coefficients (both on the total sample and on gender, age, and cultural subgroups).

This contribution is relevant for both practitioners interested in using the instrument in applied settings and researchers who want to use it in cross-cultural studies. Indeed, this new abbreviated version of the EPQ-R is a suitable and psychometrically solid instrument for the assessment of PEN-L traits among Italian and English speaking people. Another contribution concerning the cross-cultural invariance of this abbreviated version of the EPQ-R has recently been provided by McLarnon and Romero (2020), who investigated the invariance of three of the four scales (i.e., E, N, and L scales) across 35 different countries. The authors concluded that, although some differences exist among countries, the E and N scales could be optimally used when cross-cultural comparisons are of central interest.

The existing literature showed that "dark traits" such as psychoticism may influence a variety of aspects pertaining to social and individual life (e.g., Capozza et al., 2019; Colledani, 2017; Prusik & Szulawski, 2019; Shih et al., 2019). In the present study, psychoticism has been found to be correlated with negative psychosocial and behavioral outcomes, demonstrating its influential role in the life of people. However, the P scale remains rather problematic for reliability. This scale has always been recognized as the psychometrically weakest of the instrument. Future studies could be aimed at developing a new pool of items more suitable for the assessment of this trait. A limitation of this work is due to the fact that validity was investigated using self-report measures only. Future studies should try to replicate these findings using more objective behavioral measures.

REFERENCES

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*, Article 16. https://doi.org/10.1186/s12955-014-0197-x

Bobbio, A., & Manganelli, A. M. (2011). Measuring social desirability responding. A short version of Paulhus' BIDR 6. *TPM – Testing, Psychometrics Methodology in Applied Psychology*, *18*, 117-135. https://doi.org/10.4473/TPM.18.2.4

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

Boyle, G. J., Matthews, G., & Saklofske, D. H. (Eds.). (2008). *The SAGE handbook of personality theory and assessment, Vol. 2. Personality measurement and testing*. Sage Publications, Inc.

Capozza, D., Colledani, D., & Falvo, R. (2019). The Dark Triad, humanity attributions, and behavioral inclinations toward outgroups. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *26*, 463-479. https://doi.org/10.4473/TPM26.3.10

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K.-M., Falk, C. F., Meade, A., Schneider, L., King, D., Liu, C.-W., & Oguzhan, O. (2018). *Package "mirt"* (Version 1.29). Retrieved from https://cran.r-project.org/web/packages/mirt/mirt.pdf

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

Colledani, D. (2017). The mediating role of values between personality and religiosity. *International Journal of Children's Spirituality*, *22*, 120-133. https://doi.org/10.1080/1364436X.2017.1310089

Colledani, D. (2018). Psychometric properties and gender invariance for the Dickman Impulsivity Inventory. *TPM – 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*, Article 1834. https://doi.org/10.3389/fpsyg.2018.01834

Colledani, D., Anselmi, P., & Robusto, E. (2019a). Development of a new abbreviated form of the Eysenck Personality Questionnaire-Revised with multidimensional item response theory. *Personality and Individual Differences*, *149*, 108-117. https://doi.org/10.1016/j.paid.2019.05.044

Colledani, D., Anselmi, P., & Robusto, E. (2019b). 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., & Camperio Ciani, A. (2021). A worldwide internet study based on Implicit Association Test revealed a higher prevalence of adult males' androphilia than ever reported before. *The Journal of Sexual Medicine*, *18*, 4-16. https://doi.org/10.1016/j.jsxm.2020.09.011

Colledani, D., Capozza, D., Falvo, R., & Di Bernardo, G. A. (2018). The work-related basic need satisfaction scale: an Italian validation. *Frontiers in Psychology*, *9*, Article 1859. https://doi.org/10.3389/fpsyg.2018.01859

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

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, S., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*, *54*, 485-489. https://doi.org/10.1016/j.paid.2012.09.022

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

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.

Karanci, A. N., Dirik, G., & Yorulmaz, O. (2007). Reliability and validity studies of Turkish translation of Eysenck Personality Questionnaire Revised-Abbreviated. *Turkish Journal of Psychiatry*, *18*, 254-261.

Kircaburun, K., & Griffiths, M. D. (2018). Instagram addiction and the Big Five of personality: The mediating role of self-liking. *Journal of Behavioral Addictions*, *7*(1), 158-170. https://doi.org/10.1556/2006.7.2018.15

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

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates, Inc.

McLarnon, M. J. W., & Romero, E. F. (2020). Cross-cultural equivalence of shortened versions of the Eysenck Personality Questionnaire: An application of the alignment method. *Personality and Individual Differences, 163*, Article 110074. https://doi.org/10.1016/j.paid.2020.110074

Moreno, E., Muñoz-Navarro, R., Medrano, L. A., González-Blanch, C., Ruiz-Rodríguez, P., Limonero, J. T., Moretti, L. S., Cano-Vindel, A., & Moriana, J. A. (2019). Factorial invariance of a computerized version of the GAD-7 across various demographic groups and over time in primary care patients. *Journal of Affective Disorders*, *252*, 114-121. https://doi.org/10.1016/j.jad.2019.04.032

Muthén, L. K., & Muthén, B. O. (2012). *Mplus Version 7 User's Guide*. Muthén & Muthén.

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

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27. https://doi.org/10.1016/j.jbef.2017.12.004

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic metaanalysis. *General Hospital Psychiatry*, *39*, 24-31. https://doi.org/10.1016/j.genhosppsych.2015.11.005

Prusik, M., & Szulawski, M. (2019). The relationship between the Dark Triad personality traits, motivation at work, and burnout among HR recruitment workers. *Frontiers in Psychology*, *10*, Article 1290. https://doi.org/10.3389/fpsyg.2019.01290

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203-212. https://doi.org/10.1016/j.jrp.2006.02.001

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

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*, 251. https://doi.org/10.3390/brainsci10040251

Shih, S. I., Chi, N. W., Wu, C. C., & Wang, S. Y. (2019). When dark meets blue: The relations between dark triad personality and depression symptoms. *Current Psychology*. Advanced online publication. https://doi.org/10.1007/s12144-019-00549-7

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Pearson.

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). Sage.