



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE CIENCIAS HUMANAS Y
SOCIALES

**Adaptación al castellano del cuestionario de
personalidad IPIP-FFI**

Autor/a: Andrés Iglesia Fernández-Tresguerres

5º Psicología y Criminología

Director/a: José Manuel Caperos Montalbán

Madrid

2019/2020

RESUMEN

El cuestionario *International Personality Item Pool Five Factor Model* (IPIP-FFM) es un test de acceso abierto diseñado por Goldberg que permite medir los cinco factores de personalidad. En el presente trabajo se ha realizado la adaptación y validación al castellano del (IPIP-FFM). Para ello en primer lugar se realizó la traducción y retro-traducción de los ítems del test; posteriormente se administró el cuestionario de forma *online* a una muestra de 234 personas, seleccionando los 35 ítems que presentaban un mejor funcionamiento en base a su homogeneidad, validez y discriminación con el resto de las escalas. Finalmente se comprobó la dimensionalidad de la versión reducida de 35 ítems del cuestionario mediante análisis factorial y se analizó la validez convergente y divergente con los correspondientes factores medidos mediante el cuestionario NEO-FFI 3r. La escala de 35 ítems presentó buenas medidas de fiabilidad. Además, la matriz de datos se ajustó satisfactoriamente a la estructura pentafactorial teórica, tanto en el análisis factorial exploratorio como en el confirmatorio. Finalmente, las correlaciones con los respectivos factores del NEO-FFI muestran una adecuada validez discriminante, aunque la validez convergente de la subescala de Amabilidad es un poco menor de lo esperado. Los resultados son prometedores e indican que el instrumento resultante puede ser usado como medida de la personalidad en un futuro, sin embargo, sería necesario confirmar la estructura factorial propuesta en una muestra independiente.

Palabras clave

IPIP, Personalidad, Test, Validación, IPIP-FFM

ABSTRACT

The *International Item Pool Five Factor Model* (IPIP-FFM) is an open sourced test designed by Goldberg that allows the measuring of the five personality traits. In this paper, the IPIP-FFM was adapted and validated to a Castilian Spanish sample. For that, a translation and backwards translation of the items was done; Later, the test was administrated to a sample of 234 individuals, selecting the 35 items that presented the best performance, based on the item-rest correlation, validity and discrimination with the rest of the scales. Finally, the dimensionality of the reduced 35 items version of the test was confirmed through an exploratory factorial analysis and the convergent and discriminant validity with the different factors was analysed through the NEO-FFI 3r. The 35-item scale presented good reliability measures. Further, the data matrix adjusted to the theoretical five factor structure both in the exploratory and confirmatory factor analyses. Finally, the correlations with the NEO-FFI show an adequate discriminant validity, although convergent validity of the agreeableness subscale was a bit lower than expected. The results are promising and indicate that the resulting instrument may be used as a measure of personality in the near future. However, the factorial structure needs to be confirmed on an independent sample.

Keywords

IPIP, Personality, Test, Validation, IPIP-FFM

INDEX

METHOD	7
Sample	7
Instruments.....	8
IPIP-FFM.	8
NEO Five Factor Inventory revised version, 3d edition.	8
Procedure	8
Translation procedure.....	8
Test administration.....	9
Data analysis	9
RESULTS.....	11
Reliability and item functioning	11
Validity evidence based on internal structure.....	16
Exploratory factor analysis.....	16
Convergent and discriminant validity.....	24
DISCUSSION.....	24
CONCLUSIONS	27
REFERENCES	27

The study of personality has been on the rise over these past decades (Weiner & Greener, 2017). Such raise is not fortuitous, given the influence personality has on many different areas (Maples-Keller et al, 2017). Indeed, personality is being used in areas far apart from each other, such as risky driving behaviour (Akbari et al, 2019), workplace deviance (Pletzer et al, 2019) or even artificial intelligence and robotics (Craenen et al, 2018).

Although there has been a plethora of different models regarding personality, such as Eysenck's (2006) or Cattell's (Cattell, Cattell and Cattell, 1994), the Five Factor Model (FFM) is the most used one to this date (McCrae, 2009). Questioned though as it might have been in the past, this model has managed to resist such criticism proving its universality, and the stability of its traits (Deary, 2009). The Five Factor Model approaches personality as if made up by five independent factors (McCrae, 2009). Such factors are Neuroticism, Extraversion, Openness to experience, Agreeableness and Conscientiousness, and combine to account for variations in personality (Deary, 2009), thus providing '*a structure in which personality can be classified*' (McCrae, 2009).

Many different instruments have been created to measure personality, such as the Personality Assessment Inventory (Morey, 2007), or the Millon Clinical Multiaxial Inventory (Millon, Grossman & Millon, 2015). However, after the rise of the Five Factor Model (Goldberg, 1990), the most used instrument to do so is the NEO PI-R (Costa and McCrae, 1992). Created by McCrae and Costa, this instrument is comprised of 240 items, and divides each factor into 6 additional sub-factors (1992). Additionally, a shorter version of this test exists, the FFI, comprised of 60 items instead (Costa and McCrae, 1992). Through the use of these inventories, the Five Factor Model has had a notorious impact on the study of personality (McCrae, 2009).

Unfortunately, however, the study of personality comes with quite the literal cost, as the NEO PI-R, as well as a number of other validated and well-known personality tests are proprietary and thus pay-to-use (Goldberg, 1999). Therefore, this can have detrimental consequences to investigative capabilities, as researchers are required to gather substantial samples to ensure statistical correctness, which can prove to be cost prohibitive (Maples-Keller et al, 2017). Whilst economic matters are less of a hindrance nowadays, this issue still remains problematic (Weiner & Greener, 2017). Further, these questionnaires often lack in adaptability (as researchers may not remove or change items from the test), which can lead to them ending up "obsolete" over time, as they often are less likely to be revised (Goldberg, 2006). This has influenced the development of personality science, which has been, according to Goldberg '*dismally slow*' (1999).

Table 1

Factor definitions following McCrae and Sutin (2007, p. 424).

Factor	Definition: The tendency to...
Neuroticism	Experience many forms of emotional distress, have unrealistic ideas and troublesome urges. Individuals low in Neuroticism are emotionally stable, do not get upset easily, and are not prone to depression.
Extraversion	Prefer intense and frequent interpersonal interactions; be energized and optimistic. Individuals low in Extraversion are reserved and tend to prefer a few close friends to large groups of people.
Openness to experience	Seek out new experience and have a fluid style of thought. Individuals low in Openness are traditional, conservative, and prefer familiarity to novelty.
Agreeableness	Regard others with sympathy and act unselfishly. Individuals low in Agreeableness are not concerned with other people and tend to be antagonistic and hostile.
Conscientiousness	Control one's behaviour in the service of one's goals. Individuals low in Conscientiousness have a hard time keeping to a schedule, are disorganized, and can be unreliable.

In an attempt to solve some of these issues, Goldberg and colleagues formed the International Personality Item Pool (Goldberg, 1990). This collection of items is fine-tuned to measure different facets of personality but, most importantly, is free to use and modify by fellow researchers (Goldberg et al, 2006). As a result of this effort, many different instruments have been created using items from the platform, such as the IPIP-FFM, the IPIP-NEO (Goldberg, 1999) or the IPIP-NEO-60 (Maples-Keller, 2017), to name a few. Therefore, the usefulness and impact of the IPIP cannot be denied (Goldberg et al, 2006).

As it happens, validation of IPIP-based questionnaires is not foreign to Spanish speaking countries. Indeed, Cubani has made an outstanding effort validating different IPIP-based instruments in various Latin American samples (Gross & Cupani, 2016; Cupani, Pilatti, Urrizaga, Chincolla & Richaud, 2014; Cupani, 2009; Pérez, Cupani &

Beltramino, 2004), and so have other authors (Martínez-Molina & Arias, 2018). One such instrument, the IPIP-FFM (Goldberg, 1999), is comprised of fifty questions measuring personality based on the FFM. It also presents good psychometric qualities, such as a good reliability (Cronbach's alpha = .77 - .84) or a good temporal stability of the measures ($r = .71 - .84$; Cupani, 2009), which makes it a rather attractive test to use.

It should be noted, however, that one flaw exists on the aforementioned test. Due to its length and composition, whilst it does provide a general measure of the Five Factor Model, it doesn't do it in a level of depth akin to others such as the NEO PI-R, and thus cannot discriminate the different sub-factors that make up the correspondent five factors (Goldberg, 1999). Indeed, as stated by Goldberg (1999), measuring only five factors cannot discriminate the variance associated with each specific sub-factor. We find ourselves then at a trade-off between discrimination and instrument length and overall practicality. Considering how long tests as well as time constraints influence the results through the effects of fatigue (Maples-Keller et al, 2017), such trade-off is considered necessary. Overall, the IPIP-FFM is found to be a rather useful instrument, as it provides a short yet reliable measure of personality. However, in order to reduce the effects of said fatigue and attempting to obtain a truly practical and quick test, we attempted to create a test based on the IPIP-FFM, with seven items per factor. Although it is true that good internal consistency can be obtained with at least four items (Harvey, Billings & Nillan, 1985), and that a minimum of three items may provide a valid measure of the sought construct (Hair et al, 2010), when it comes to the distribution of the items in their proper factors, it has been found that the more items measuring the same factor, said factor will be more stable and its measure will be much more precise (Ferrando & Anguiano, 2010). Therefore, seven items seems a good compromise between administration efficiency and psychometric qualities.

Alas, there is a surprising lack of Castilian Spanish validation of any IPIP based instrument. Therefore, it is the objective of this paper to validate a 35-items version of the IPIP-FFM based test to the Castilian Spanish sample. Such efforts are intended as an entryway for further validation of similar questionnaires on Castilian Spanish samples.

METHOD

Sample

The sample consisted of 234 valid participants. Said sample was composed of 130 males (55,56%), 101 females (43,16%) and two individuals (0.01%) who decided not to reveal

their gender. Their age ranged from 18 years (minimum age to partake in the study) and 85 years (mean = 39.15; SD = 14.74). As for educational level, 22 participants (9.4%) had at least a doctorate, 60 (25.64%) had a masters degree, 93 (39.64) had gone to the university, 50 (21.36%) had finished their baccalaureate, 5 (2.13%) had attended professional education, 3 (1.28%) had finished their secondary education whilst lastly one individual (0.42%) only finished their primary school. All of the participants had lived most of their life in Spain. Several individuals were excluded of this sample due to selection criteria, in particular, being underaged (3 participants) and not being a Spanish national (16 participants).

Instruments

IPIP-FFM (Goldberg, 1999). The test is comprised of fifty items and may be freely downloaded from the internet. The items are divided in five different factors, following the structure of the Big Five model. In particular, the test is divided in agreeableness (A), extraversion (E), intellect (I) emotional stability (E), and conscientiousness (C). The items ask the participant to grade in a Likert scale how accurately the item describes them (with the scale ranging from 1 -very inaccurate- to 5 -very accurate-). Goldberg (1999) reports adequate psychometric properties, such as Cronbach's alpha = .77 - .84.

NEO Five Factor Inventory revised version, 3d edition (NEO FFI-R; Spanish version; Cordero, Pamos & Seisdedos, 2008). This inventory has 60 items, distributed in the five factors of the Big Five Model (with 12 items per factor): extraversion (E), agreeableness (A), emotional stability (E), intellect (I) and conscientiousness (C). Participants are asked to indicate the degree to which the items described them, in a Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). Participants are asked to answer several control items wherein they have to mark the answer that the test asks. Costa and McCrae report adequate psychometric properties, such as Cronbach's α of .79 - .86. Further, test-retest reliability is reported to be of .86 - .90.

Procedure

Translation procedure. In order to validate the test to Spanish, given the lack of previous work in doing so to Castilian Spanish, the translation of the test was undertaken. Said translation followed the International test commission's (ITC) guidelines for translating and adapting test (Second edition; ITC, 2017).

Firstly, the fifty items of the test were translated to Spanish. Additionally, the translated tests in Latin American were gathered and organized according to the original English items. Using these, as well as the above-mentioned initial translation, a second translation was agreed upon by two writers, taking into consideration the meaning of the original wording as well as the possible psychometric influence of the different translation possibilities. Indeed, through this process literal translation was avoided, trying to instead ponder about the translation that would keep the full meaning of the item (Muñiz, Elosua & Hambleton, 2013). This second translation was then translated back to English by a bilingual speaker, to then reconcile the result with the original English version. This comparison was performed by a bilingual English professor, acquainted with American culture. As a result of the comparison several items were changed in order to ensure the quality of the translation, as seen in table 2.

Table 2
Translation reconciliation

Initial translation	Spanish-English translation	Original version	Reconciliated version
Me altero con facilidad	I get disturbed easily	Am easily disturbed	Me molesto con facilidad
Me disgusto con facilidad	I get angry easily	Get upset easily	Me altero con facilidad
Soy un desastre	I am a mess	Make a mess of things	Nada me sale bien

Test administration. The test was administered via two different methods, both using an electronic survey tool. On the one hand, a part of the sample was presented with a series of demographic questions as well as the translated test. On the other hand, the remaining sample was also asked the demographic questions and the translated test, immediately followed by the NEO-FFI questionnaire. Both sub samples were provided an email, to answer any question they might have regarding the test administration.

Data analysis

Firstly, the performance of the 50 individual items was analysed in order to select the 35 items (seven per factor), that presented better results. In order to select them, the item-rest correlation, the validity and discrimination indexes of the individual items were calculated. Indeed, as Fabrigar et al (1999, p. 283) put it, “*when validity information is available, it should be used as a basis for selecting items*”. For the item-rest correlation, a threshold of 0.35 was set, as correlations above said threshold are considered adequate (Carretero-Dios & Perez, 2005). In the case of convergent validity index we used the

correlation between the item and their corresponding factor score in the NEO-FFI; for the discriminant indexes we employed the maximum correlation between the item and the other four NEO-FFI factors; in both we used the 0.35 threshold. Finally, the mean, standard deviation and Cronbach alpha was calculated for both the original scale and the reduced 35 items scale, to evaluate the reliability of the shortened scale.

Following this analysis, an exploratory factor analysis (EFA) was undertaken to analyse the factorial structure of the resulting scale. In order to ensure the suitability of the data for such analysis, prior to the analysis itself the skewness and kurtosis of the items were considered (Carretero-Dios & Pérez, 2005). For both markers, values in between 1 and -1 for both skewness and kurtosis would be considered excellent as a measure of the data's normal distribution (George & Mallery, 2019). Further, Bartlett's Test of Sphericity and KMO measure of sampling adequacy were also calculated with overall satisfactory results. Therefore, a principal components analysis was performed on the data set. Indeed, Carretero-Dios & Pérez (2005) argue that in a sample such as the present one, wherein the number of items is higher than 20 and an adequate inter-item correlation, the differences between the factorial results of the different methods is negligible. Further, they recommend the use of the principal components, as it is easier to compute and interpret. As recommended by Carretero-Dios and Perez (2005) and Izquierdo, Olea and Abad (2014), loadings below 0.4 were dismissed. Additionally, given how personality factors correlate mildly (Cupani, 2009), a Promax rotation was used (Fabrigar, Wegener, McCallum & Strahan, 1999). Not doing so could end up with misleading results, as the model would be fitted to an unempirical relation of the factors (orthogonal), which would be untrue to how the model realistically functions (Fabrigar et al, 1999). Although the Kaiser rule is one of the most used methods for determining what factors to keep, this measure can sometimes be problematic (Fabrigar et al, 1999). Therefore, Horn's parallel analysis (1965) and the scree plot (Catell, 1966) were obtained.

Following that, a confirmatory factor analysis was also calculated, in order to test the resulting factorial structure. The size of the sample might be lacking for this kind of analysis, as argued by some scholars who demand ten individuals per analysed variable, nor those who ask for a laxer 5 individuals per variable (close though as the N, 234 would be to actually doing so; Fabrigar et al, 1999). However, Izquierdo et al. (2014), argue that with at least seven variables per factor and communalities higher than .5, a sample ranging from 100 to 200 is sufficient. As for what measures to use regarding the goodness of fit of the proposed factorial structure, the most used are the goodness of fit index (GFI), the

adjusted goodness of fit index (AFGI), the root mean square error of approximation (RMSA), the Standard Residual Mean Root (RMSR) and the non-normed fit index (NNFI; Carretero-Dios & Pérez, 2005). However, out of those, the RMSA seemed to be particularly useful in our sample scenario. Indeed, the usage of RMSEA is preferred when working with a small sample, as it is less influenced by sample size and distribution (Hu & Bentler, 1998), whereas. For this particular measure, values ranging from .05 to .08 were considered to indicate a fair fit (Fabrigar et al, 1999; Hu & Bentler, 1998). On the other hand, two indicators may be used, the CFI and TLI, which are not influenced by sample size (Fabrigar et al, 1999). For these, a threshold of .95 will be used (Fabrigar et al, 1999).

Finally, convergent and divergent validity was analysed through the correlations between the resulting test and Costa and McCrae's NEO-FFI. Such scores were calculated using the sum of the direct values of the items in their respective scale.

All the data analysis was done through the use of Jamovi (Ver. 1.1.9.0; The jamovi project, 2019). Said program is based on R, a statistical analysis based operating system (R Core Team, 2018), and allows the installation of specific packages, such as the used in these analysis, psych, which allows for the computation of reliability and factor analysis (Revelle, 2019).

RESULTS

Reliability and item functioning

According to our objective 15 items were removed from the scale, 5 from each factor (Table 3). Sometimes the decision to remove an item was straightforward, as said item had notorious deficiencies in the different markers when compared with the rest of the items of the subscale. For instance, the item *Suelo preocuparme por las cosas* showed the lowest score in both the item-rest correlation and the validity index (Table 5). However, the decision wasn't as clear in other items wherein a value was lower than the rest of the items, but the other ones were not. Whilst in general internal consistency, as measured by the item-rest correlation, was prioritized, sometimes validity was prioritized instead. Such was the case of the item *Me siento cómodo rodeado de la gente*, which, although had a lower test-rest correlation when compared to *No soy una persona habladora*, also had a much higher validity index than its comparison. Particularly difficult was the decision of what items to remove in the Intellect scale. Whilst the removal of the items *Entiendo*

rápidamente las cosas and *Estoy lleno/a de ideas* were straightforward, since they did not correlate much with the intellect scale of the NEO-FFI in a statistically significant way, the decision was harder with the remaining item. Ultimately, the decision came down to the items *Tengo un vocabulario amplio* and *Paso tiempo reflexionando en las cosas*. Indeed, as the later showed a better item-rest correlation, but a worse validity index, the former did so in reverse. The decision to remove the item was based on the discrimination index, as *Paso tiempo reflexionando en las cosas* didn't correlate with any factor outside its own, but *Tengo un vocabulario amplio* did so with extraversion.

Table 3
Removed items from each subscale

Agreeableness	Extraversion	Intellect	Emotional stability	Conscientiousness
Me interesan las personas	No me importa ser el centro de atención	Tengo un vocabulario amplio	Habitualmente estoy relajado	Presto atención a los detalles
Soy una persona sensible	No soy una persona habladora	Entiendo rápidamente las cosas	Suelo preocuparme por las cosas	Dejo mis cosas en cualquier sitio
Hago que la gente se sienta a gusto	Tengo poco que decir	Estoy lleno/a de ideas	Me molesto con facilidad	Nada me sale bien

There was a logical reduction in the reduced test reliability, as measured by the subscales as a direct result of the removal of the items. However, while the alphas are indeed reduced, their values still remain near .7, ranging from .695 to .867, with a median of .770.

Table 4
Reliability statistics per subscale in the full and reduced version of the IPIC

Factor	Full test			Reduced test		
	Mean	SD	Cronbach's α	Mean	SD	Cronbach's α
Agreeableness	4.07	0.421	.763	4.04	0.473	.751
Extraversion	3.20	0.627	.838	3.11	0.649	.806
Intellect	3.57	0.468	.762	3.64	0.524	.695
Emotional stability	3.29	0.697	.874	3.46	0.785	.867
Conscientiousness	3.63	0.523	.788	3.79	0.607	.770

Table 5

IPIP item statistics (reliability and validity indexes)

Item	Mean	Standard deviation	Item-rest correlation	Validity index	Discrimination index
<i>Agreeableness</i>					
Me interesan las personas ^a	4.43	0.834	.375	.226	.346
Me preocupan los sentimientos de los demás	4.32	0.653	.663	.343	.208
Soy una persona sensible ^a	3.91	0.727	.323	.123	.132
Dedico tiempo a los demás	3.79	0.700	.500	.262	.366
Empatizo con las emociones de los otros	4.13	0.715	.556	.315	.338
Hago que la gente se sienta a gusto ^a	4.07	0.635	.362	.133	.338
Me preocupo poco por los demás	2.84	0.599	.578	-.132	-.345
Soy brusco tratando a otras personas	4.09	0.858	.572	-.390	-.237
No me interesan los problemas de los demás	4.28	0.755	.491	-.235	-.236
No estoy muy interesado en la gente	4.24	0.824	.604	-.305	-.322
<i>Extraversion</i>					
Soy el alma de la fiesta	3.01	0.775	.557	.538	-.179
Me siento cómodo rodeado de gente	3.09	0.904	.481	.598	-.194
Suelo empezar las conversaciones	2.69	0.942	.669	.524	-.232
En las fiestas, hablo con muchas personas distintas	2.63	1.026	.631	.579	-.238

No me importa ser el centro de atención ^a	2.82	1.069	0.388	.326	-.129
No soy una persona habladora ^a	2.68	1.164	.582	-.407	-.108
Suelo quedarme en un segundo plano	3.21	0.806	.550	-.462	.242
Tengo poco que decir ^a	2.19	0.883	.531	-.482	.351
No me gusta acaparar la atención	3.29	1.011	.473	-.484	.208
Soy callado con los desconocidos	2.79	1.136	.626	-.522	.264

Intellect

Tengo un vocabulario amplio ^a	2.63	0.993	.5672	.348	-.312
Tengo una imaginación viva	2.59	0.998	.5278	.330	-.277
Tengo excelentes ideas	2.95	1.206	.6088	.347	.271
Entiendo rápidamente las cosas ^a	4.10	0.662	.0367	.169	.440
Uso palabras complejas	2.44	0.984	.6139	.416	.119
Paso tiempo reflexionando en las cosas	2.46	0.993	.7576	.312	.096
Estoy lleno/a de ideas ^a	2.27	0.916	.7381	.203	.139
Me cuesta entender los conceptos abstractos	2.15	1.012	.6147	-.305	.343
No estoy interesado en las ideas abstractas	2.12	1.025	.7232	-.538	-.218
No tengo una buena imaginación	1.97	0.903	.6384	-.370	.323

Emotional stability

Habitualmente estoy relajado/a ^a	3.81	0.723	.372	-.458	.231
Es raro que me sienta triste	3.88	0.854	.573	-.635	.368

Me estreso con facilidad	3.45	0.733	.527	.549	-.331
Suelo preocuparme por las cosas ^a	3.90	0.725	.459	-.0240	.281
Me molesto con facilidad ^a	2.94	0.861	.304	.375	-.377
Me altero con facilidad	3.80	0.658	.343	.515	-.361
Mi estado de ánimo fluctúa mucho	2.46	0.641	.420	.518	-.421
Tengo cambios de humor frecuentemente	3.69	0.975	.450	.607	-.428
Me irrito con facilidad	2.67	0.885	.549	.570	-.407
Me siento triste a menudo	2.88	0.935	.539	.797	-.658

Conscientiousness

Siempre estoy listo	3.75	0.797	.63	.648	-.539
Presto atención a los detalles ^a	4.11	0.782	.244	.195	.314
No dejo las cosas para más tarde	3.14	1.115	.647	.607	-.339
Me gusta que las cosas estén ordenadas	3.91	0.955	.588	.494	-.201
Sigo los horarios	3.65	0.99	.564	.492	-.207
Soy exigente en mi trabajo	4.19	0.701	.498	.601	-.463
Dejo mis cosas en cualquier sitio ^a	2.47	0.745	.612	-.450	.356
Nada me sale bien ^a	3.31	0.862	.410	-.519	.557
A menudo se me olvida poner las cosas en su sitio	3.72	1.001	.675	-.563	.476
Evito mis responsabilidades	4.32	0.826	.626	-.658	.575

Note. ^a = removed item.

Internal structure

Exploratory factor analysis

Bartlett's Test of Sphericity yielded a value of $X^2(630) = 3310$; $p < .001$, whereas the KMO measure of sampling adequacy yielded a value of .796. Horn's parallel analysis revealed a five-factor structure. Similar results can be seen on the scree plot (figure 1).

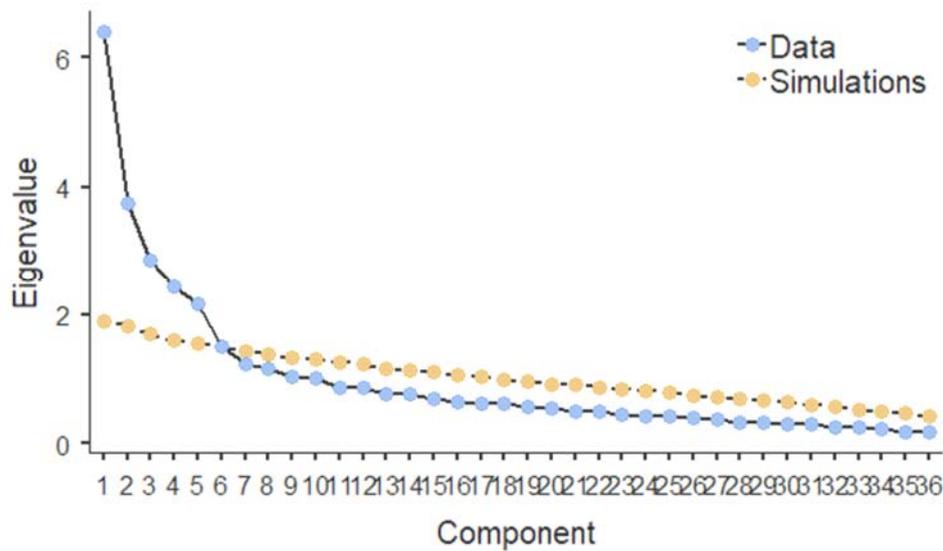


Figure 1: Scree plot

Indeed, a significant drop in the scree plot past the fifth factor, which implies a significant reduction of the variability explained by the model should additional factors be considered. On the other hand, as seen on table 6, all factors saturated accordingly in their theoretical factor. That said, two items saturated under the .4 threshold (*Me preocupo poco por los demás* and *Siempre estoy listo*). The first factor explained 14.2% of the variance, with the rest of the factors explaining 10.21%, 8.57%, 8.22% and 7.22% respectively. The total variance explained by the model was 48.9%.

Table 6
Component Loadings

	Component					Uniqueness
	1	2	3	4	5	
<i>Agreeableness</i>						
Me preocupan los sentimientos de los demás			.734			.453
Dedico tiempo a los demás			.516			.570
Empatizo con las emociones de los otros			.670			.514
Me preocupo poco por los demás			-.395			.706
Soy brusco tratando a otras personas			-.554			.527
No me interesan los problemas de los demás			-.653			.545
No estoy muy interesado en la gente			-.656			.473
<i>Extroversion</i>						
Soy el alma de la fiesta		.678				.484
Me siento cómodo rodeado de gente		.700				.480
Suelo empezar las conversaciones		.632				.525
En las fiestas, hablo con muchas personas distintas		.804				.395
Suelo quedarme en un segundo plano		-.488				.697
No me gusta acaparar la atención		-.504				.620
Soy callado con los desconocidos		-.715				.443

Intellect

Tengo una imaginación viva	.621	.504
Tengo excelentes ideas	.642	.528
Uso palabras complejas	.472	.775
Paso tiempo reflexionando en las cosas	.463	.716
Me cuesta entender los conceptos abstractos	-.607	.605
No estoy interesado en las ideas abstractas	-.693	.482
No tengo una buena imaginación	-.552	.610

Emotional stability

Es raro que me sienta triste	-.643	.449
Me estreso con facilidad	.765	.446
Me altero con facilidad	.893	.270
Mi estado de ánimo fluctúa mucho	.469	.723
Tengo cambios de humor frecuentemente	.752	.356
Me irrito con facilidad	.858	.263
Me siento triste a menudo	.734	.328
Me molesto con facilidad	.804	.388

Conscientiousness

Siempre estoy listo	.396	.592
No dejo las cosas para más tarde	.784	.422

Me gusta que las cosas estén ordenadas	.766	.443
Sigo los horarios	.738	.484
Soy exigente en mi trabajo	.445	.608
A menudo se me olvida poner las cosas en su sitio	-.655	.472
Evito mis responsabilidades	-.522	.489

Note. 'Promax' rotation was used

Additionally, the factors did not correlate strongly with each other. Indeed, the higher correlations found were Extraversion-Intellect ($r = -.255$), Conscientiousness-Agreeableness ($r = .239$) and Extraversion-Agreeableness ($r = .237$).

Confirmatory factor analysis

As for the CFA, all the items saturated in their factors significantly. Regarding the goodness of fit, the χ^2 showed statistical significance, whilst the RMSEA was contained within the range previously set to consider it to show an adequate fit. That said, both the CFI and the TLI didn't exceed the 0.95 threshold (See table 8).

Table 7

Fit Measures

Test for Exact Fit					RMSEA 90% CI		
χ^2	df	P	CFI	TLI	RMSEA	Lower	Upper
1250	550	< .001	.742	.721	.0738	.0683	.0792

Finally, regarding the factor correlations, a similar correlation pattern found in the exploratory factor analysis was found. Indeed, both Intellect ($r = .340$; $p < .001$) and Agreeableness ($r = .400$; $p < .001$) correlated in a statistically significant way and Extraversion respectively. However, Emotional stability also correlated with Conscientiousness ($r = .423$; $p < .001$) in a statistically significant way.

Table 8

Correlation matrix between factor in the exploratory (above the diagonal) and the confirmatory analyses (below the diagonal).

	Emotional stability	Extraversion	Conscientiousness	Agreeableness	Intellect
Emotional stability	—	-.190	-.223	-.139	-.097
Extraversion	.203	—	.053	.237	.256
Conscientiousness	-.423	.115	—	.239	.055
Agreeableness	-.248	.400	.239	—	.126
Intellect	-.176	.340	.105	.007	—

Table 9
Factor loadings

Item	Estimate	SE	Z	p	Stand. Estimate
<i>Agreeableness</i>					
Me preocupan los sentimientos de los demás	.379	0.0422	8.99	< .001	.608
Dedico tiempo a los demás	.417	0.0478	8.73	< .001	.588
Empatizo con las emociones de los otros	.327	0.0481	6.80	< .001	.482
Me preocupo poco por los demás	-.305	0.0497	-6.14	< .001	-.432
Soy brusco tratando a otras personas	-.378	0.0640	-5.91	< .001	-.424
No me interesan los problemas de los demás	-.456	0.0483	-9.45	< .001	-.629
No estoy muy interesado en la gente	-.606	0.0554	-10.93	< .001	-.712
<i>Extraversion</i>					
Soy el alma de la fiesta	.546	0.0525	10.41	< .001	.660
Me siento cómodo rodeado de gente	.588	0.0591	9.95	< .001	.634
Suelo empezar las conversaciones	.621	0.0577	10.77	< .001	.675
En las fiestas, hablo con muchas personas distintas	.763	0.0649	11.75	< .001	.722
Suelo quedarme en un segundo plano	-.358	0.0554	-6.45	< .001	-.441
No me gusta acaparar la atención	-.451	0.0668	-6.76	< .001	-.458
Soy callado con los desconocidos	-.773	0.0687	-11.25	< .001	-.701
<i>Intellect</i>					

Tengo una imaginación viva	.778	0.0561	13.88	< .001	.858
Tengo excelentes ideas	.413	0.0537	7.69	< .001	.522
Uso palabras complejas	.160	0.0655	2.45	.014	.178
Paso tiempo reflexionando en las cosas	.174	0.0536	3.24	.001	.231
Me cuesta entender los conceptos abstractos	-.204	0.0694	-2.94	.003	-.220
No estoy interesado en las ideas abstractas	-.333	0.0670	-4.98	< .001	-.360
No tengo una buena imaginación	-.705	0.0568	-12.41	< .001	-.758
<i>Emotional stability</i>					
Es raro que me sienta triste	.663	0.0632	10.48	< .001	.0647
Me estreso con facilidad	-.782	0.0726	-10.77	< .001	-.656
Me altero con facilidad	-.795	0.0600	-13.25	< .001	-.768
Mi estado de ánimo fluctúa mucho	-.419	0.0644	-6.50	< .001	-.428
Tengo cambios de humor frecuentemente	-.826	0.0595	-13.88	< .001	-.788
Me irrita con facilidad	-.868	0.0584	-14.85	< .001	-.827
Me siento triste a menudo	-.776	0.0583	-13.32	< .001	-.772
<i>Conscientiousness</i>					
Siempre estoy listo	.388	0.0583	6.64	< .001	.477
No dejo las cosas para más tarde	.716	0.0757	9.46	< .001	.633
Me gusta que las cosas estén ordenadas	.518	0.0671	7.72	< .001	.569
Sigo los horarios	.484	0.0685	7.07	< .001	.498

A menudo se me olvida poner las cosas en su sitio	-.689	0.0744	-9.26	< .001	-.647
Evito mis responsabilidades	-.575	0.0622	-9.25	< .001	-.655
Soy exigente en mi trabajo	.352	0.0498	7.07	< .001	.503

Convergent and discriminant validity

As seen in table 9, all the factors mainly correlate in a statistically significant way with their respective factor in the NEO-FFI test. The correlations of the factors with their equivalents ranged from .417 to .798, with a median of .765. There were other correlations of note. Agreeableness correlated with both IPIP and NEO’s extraversion. Intellect did the same with NEO’s Extraversion and NEO’s Neuroticism. Extraversion correlated with NEO’s neuroticism and finally Conscientiousness correlated with both NEO’s Extraversion and NEO’s Neuroticism.

Finally, it is also of note that there were some correlations between the NEO factors. Indeed, Neuroticism correlated with both Extraversion (-.456) and Conscientiousness (-.618) and Extraversion correlated with Conscientiousness (.360).

Table 10
IPIP-NEO correlations

NEO		Agreeableness	Extraversion	Intellect	Emotional stability	Conscientiousness
Agreeableness	<i>r</i>	.417	-.013	.042	.320	-.025
	<i>p-value</i>	< .001	.889	.640	< .001	.783
Extraversion	<i>r</i>	.385	.749	.321	.302	.228
	<i>p-value</i>	< .001	< .001	< .001	< .001	.010
Intellect	<i>r</i>	.188	.061	.605	.111	.057
	<i>p-value</i>	.035	.497	< .001	.214	.530
Neuroticism	<i>r</i>	-.231	-.316	-.362	-.762	-.476
	<i>p-value</i>	.009	< .001	< .001	< .001	< .001
Conscientiousness	<i>r</i>	.220	.210	.251	.462	.760
	<i>p-value</i>	.013	.018	.005	< .001	< .001

DISCUSSION

Overall, the results of the analysis were rather satisfactory. Indeed, the items presented a good reliability, which, although obviously reduced, was on par with the translated 50 items scale. Further, the alpha values were also similar to those found by other validations of the test (Goldberg, 1999, Cupani, 2009; Gross & Cupani, 2016). An exception to the rule was the Intellect factor, which presented an alpha of .695.

Satisfactory results were also obtained in the factorial analysis, as they revealed a solid five factor structure that matched the Big Five model. Certainly, all the items distributed themselves according to said theoretical framework, with each of the items

saturating in their respective factors. However, a couple of the items did not do so strongly enough, as they didn't saturate over the established .4 threshold. That said, considering the saturation are .396 and .395, they may be deemed as proper saturations on the factor.

Less clear were the results of the confirmatory factor analysis. Whilst it is true that the RMSEA was contained within the .05 - .08 range to be considered a fair fit (Fabrigar et al, 1999; Hu & Bentler, 1998), the test for exact fit was statistically significant and neither the TLI nor the CFI were sufficient. Additionally, whilst still near statistical significance, some of the items in the Intellect factors saturated somewhat poorly with values under .3. It might be argued, therefore that the insufficiency of the TLI and CFI was due to model misspecification, as these markers have shown to be influenced moderately by it (Fabrigar et al, 1999). However, it is more likely that these results were due to the reduced size of the sample.

That said, these results may also be a manifestation of a different distribution of the underlying factors. Whilst in the final factor analysis the items distributed in a five-factor structure, the Intellect factor showed a tendency to split into two separate factors. The separation of the intellect factor has been previously reported in previous studies (Cupani, 2009; Ashton et al, 2004). Furthermore, there are colleges who claim that the Intellect factor is indeed made up of two separate factors (García, Aluja, García & Cuevas, 2005; Nausbaum & Silvia, 2011), or even four (Kaufman, 2013). In the more common split into two factors, the two facets of the intellect factor are considered as separate factors, splitting openness from intellect, with the former being related to creativity and the latter being related to fluid intelligence (Nausbaum & Silvia, 2011). Indeed, this separation was apparently replicated in our study, as the items that split in either factor seemed to follow this pattern, as one set of items referred to creativity (such as *no tengo una buena imaginación*), whereas the second one might refer to fluid intelligence (such as *no me gustan los conceptos abstractos*).

Table 11
Split Intellect items distribution

Factor 1	Factor 2
Tengo un vocabulario amplio	Tengo una imaginación viva
Uso palabras complejas	Tengo excelentes ideas
Paso tiempo reflexionando en las cosas	No tengo buena imaginación
No estoy interesado en ideas abstractas	

Nonetheless, these results provide sound evidence of the factorial validity of the translated test. Further, evidences of validity may also be found on the correlation between the NEO-FFI and reduced IPIP-FFI scores. Truly, the five factors of the test correlated accordingly with their respective NEO factors. That said, although the variance of the scales might be explained solely by one the five personality factors each, both the Agreeableness and Intellect scales casted a small shadow of doubt. This is not unheard of, as previous adaptations of the test have reported lower correlations with their respective scale, as happens in our sample (Gow, Whiteman, Pattie & Deary, 2005). Indeed, although the factors correlated with their NEO equivalents in a statistically significant way, they did so with other factors to a lesser extent. Both factors correlated with NEO's Extraversion, with Intellect also correlating with NEO's Neuroticism. This correlation, whilst not particularly strong might help explain why the correlations with their equivalent NEO factor is smaller than in the other three factors. Certainly, some items might be measuring aspects of another factor instead of the factor they were supposed to measure (or both at the same time). However, it should also be noted that studies have shown that the intellect subfactor correlates with extraversion, significantly so, with the subfacet of sensation seeking (García, Aluja, García & Cuevas, 2005). Further, the correlations between the factors is generally low. Therefore, it is safe to assume that the scale presents a good discriminant validity.

Speaking about the resulting scale as a whole, in general the selected items seem appropriate, formally speaking. In that sense, there's an appropriate distribution of regular and inverse items, as they generally distribute themselves in a 4:3 ratio. However, that doesn't happen in the Emotional stability subscale, where the ratio in said subscale is 1:6, although in the original scale the ratio is 2:8. Further iterations of this test might want to aim for a balance between positively and negatively worded items.

Furthermore, in general there seems to be no redundant items, as most of them were actually removed through happenstance during the process. That said, items *me altero con facilidad* and *me irrito con facilidad* on the one hand and *me cuesta entender los conceptos abstractos* and *no estoy interesado en conceptos abstractos* on the other one might be considered so, and as such might need some rework.

However, it should be noted that these results are provisional, until further research confirms or deny them. Indeed, lacking a big enough sample, the CFA was performed without optimal conditions. Further, considering that the EFA and CFA were

done on the same sample, there is a risk of overfitting the results to the sample. Therefore, further investigation must be done in order to ensure the replicability of the results.

CONCLUSIONS

Throughout this paper, we attempted to validate a short IPIP-FFI based to a Castilian Spanish environment. Firstly, the test was translated into Spanish, taking maximum care to ensure the quality of the translation through a backwards translation method. After analysing the performance of the individual items, 35 were selected to make up the final scale. Then the reliability of said scale was analysed. The factorial analysis of the collected data followed a five-factor structure, as expected by the theoretical basis of the test (Goldberg, 1999). However, whilst the exploratory factor analysis revealed a good factorial structure, some problems arose during the confirmatory factor analysis.

Overall, the results are rather satisfactory and prove that the current translated version of the IPIP-FFI test might be a good start towards an adequate measure of personality under the Big Fiver personality model. However, the Intellect and Agreeableness subfactor presented inadequate results, as their correlations with their respective factors didn't surpass 0.7. Therefore, the items of said sub scales need be revised in order to ensure adequate psychometric properties. Further investigation should also attempt to gather more sample, in order to, not only ensure that the conclusions herein contained are replicated, but also to allow for a solid confirmatory factor analysis. In that sense, it might be wise to collect sample using the fifty translated items, in order to select from said pool the items that work better. Finally, temporal stability of the measures should also be considered.

REFERENCES

- Akbari, M. et al. (2019). Meta-analysis of the correlation between personality characteristics and risky driving behaviours. *Journal of injury and violence research*, 11(2), 107-122. doi: 10.5249/jivr.v11i2.1172
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356–366. doi: [10.1037/0022-3514.86.2.356](https://doi.org/10.1037/0022-3514.86.2.356)

- Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R. Cattell (Ed.): *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.
- Cattell, R. B., Cattell, A. K. and Cattell, H. E. P. 1994. *The Sixteen Personality Factor Questionnaire*, 5th edn. Champaign, IL: Institute for Personality and Ability Testing.
- Carretero-Dios, H. & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5(3), 521-551.
- Craenen, B., Deshmukh, A., Foster, M. E., and Vinciarelli, A. (2018). Do we really like robots that match our personality? the case of big-five traits, god-speed scores and robotic gestures. *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Doi: 10.1109/ROMAN.2018.852672
- Cook, J., Hepworth, S., Wall, T. & Warr, P. (1981). *The experience of work: A compendium and review of measures and their use*. London: Academic Press.
- Cordero, A., Pamos, A. & Seisdedos, N. (2008). *Inventario neo reducido de cinco factores (NEO-FFI), Manual profesional (3º edición revisada y ampliada)*. Madrid: TEA ediciones.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Cupani, M. (2009). El cuestionario de personalidad ipip-ffm: Resultados preliminares de una adaptación en una muestra de preadolescentes argentinos. *Perspectivas en psicología*. 6, 51-58.
- Cupani, M., Pilatti, A., Urrizaga, A., Chincolla, A. & Richaud, M. C. (2014). Inventario de personalidad IPIP-NEO: estudios preliminares de adaptación al español en estudiantes argentinos. *Revista mexicana de investigación en psicología*. 6(1), 55-73.
- Deary, I. J. (2009). The trait approach to personality. In Corr, J. P. & Matthews, G. (Eds.), *The Cambridge Handbook of personality psychology*. (pp. 89-109). New York, United States of America: Cambridge University Press.
- Elosua, P. (2003). Sobre la validez de los test. *Psicothema*, 15(2), 315-321.
- Eysenck, H. J. (2006). *The biological basis of personality*. New York: Transaction Publishers.
- Fabrigar, L., Wegener, D., MacCallum, Robert & Strahan, Erin. (1999). Evaluating the use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, 4(3), 272-299.

- Ferrando, P. & Anguiano, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del psicólogo*, 31(1), 18-33.
- García, L., Aluja, A., García, O., & Cuevas, L. (2005). Is openness to experience an independent personality dimension? Convergent and discriminant validity of the openness domain and its NEO-PI-R Facets. *Journal of individual differences*, 26, 132-138.
- George, D. & Mallery, D. *IBM Statistics 25 Step by Step A simple guide reference*. New York, NY: Routledge
- Goldberg, L. R. (1999). A broad bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five Factors models. *Personality psychology in Europe*, 7, 7-28.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.
- Gow, A., Whiteman, C., Pattie, A. & Deary I. (2005). Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and individual differences*, 39(2), 317-329.
- Gross, M. N. & Cupani, M. (2016). Adaptación de los 100 ítems IPIP para medir los cinco grandes factores. *Revista Mexicana de Psicología*, 33, 17-29.
- Harvey, R. J., Billings, R. S., & Nillan, K. J. (1985). Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *Journal of Applied Psychology*, 70(3), 461-468.
- Hair, J., Black, W., Babin, B. & Anderson, R. (2010). *Multivariate data analysis*. Harlow, Essex: Pearson education limited.
- Horn, J. (1965). A rationale and test for the number of factor in factor analysis. *Psychometrika*, 30, 179-185.
- Hu, L & Bentler, P., (1998). Fit Indices in Covariance Structure Modelling: Sensitivity to Underparameterized Model Misspecification. *Psychological Methods*, 4(3), 424-453.
- Izquierdo, I., Olea, J. & Abad, J., (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, 26(3), 395-400.

- Kaufman, S. (2013). Opening up Openness to Experience: A Four-Factor Model and Relations to Creative Achievement in the Arts and Sciences. *Journal of creative behaviour*, 47, 233-255.
- Pletzer, J., Bentvelzen, M., Oostrom, J., & de Vries, R. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behaviour*, 112, 369-383. Doi: <https://doi.org/10.1016/j.jvb.2019.04.004>
- Maples-Keller, J. L., Williamson R. L., Sleep C. E., Carter N. T., Campbell W. K. & Miller J. D. (2017). Using Item Response Theory to Develop a 60-Item Representation of the NEO PI-R Using the International Personality Item Pool: Development of the IPIP-NEO-60, *Journal of Personality Assessment*, 1, 4-15. Doi: 10.1080/00223891.2017.1381968
- Martínez-Molina, A. & Arias, V. B. (2018). Balanced and positively worded personality short-forms: Mini-IPIP validity and cross-cultural invariance. *PeerJ* 6:e5542 <https://doi.org/10.7717/peerj.5542>
- Muñiz, J. & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 1, 7-16.
- Muñiz, J., Elosua, P. & Hambleton, R. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25(2), 151-157.
- McCrae, R. R. The Five-Factor Model of personality traits: consensus and controversy. In Corr, J. P. & Matthews, G. (Eds.), *The Cambridge Handbook of personality psychology*. (pp. 148-161). New York, United States of America: Cambridge University Press.
- Millon, T., Grossman, S., & Millon, C. (2015). *MCMI-IV: Millon Clinical Multiaxial Inventory IV Manual*. Bloomington, MN: NCS Pearson.
- Morey, L. C. (2007). *Personality Assessment Inventory professional manual (2nd ed.)*. Odessa, FL: Psychological Assessment Resources.
- Nusbaum, E. & Silvia, P. (2011). Are Openness and Intellect distinct aspects of Openness to Experience? A test of the O/I model. *Personality and individual differences*, 52(5), 571-574.

- Pérez, E., Cupani, M. & Beltramino, C. (2004). Adaptación del Inventario de Personalidad 16PF-IPIP a un Contexto de Orientación. Estudio Preliminar. *Evaluar*, 4, 23-49.
- R Core Team (2018). *R: A Language and environment for statistical computing*. [Computer software]. Retrieved from <https://cran.r-project.org/>.
- Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. [R package]. Retrieved from <https://cran.r-project.org/package=psych>.
- The jamovi project (2019). *jamovi*. (Version 1.1) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- Weiner, I. B. Greene, R. L. (2017). *Handbook of personality assessment*, Hoboken, New Jersey, United States of America: John Wiley & Sons.