Psychiatry Research 304 (2021) 114134

Contents lists available at ScienceDirect

Psychiatry Research

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

Cross-Cultural Measurement Invariance in the Personality Inventory for DSM-5[☆]

M.A. Sorrel[a,*], L.F. García[a,b], A. Aluja[b,c], J.P. Rolland[d], J. Rossier[e], I. Roskam[f], F.J. Abad[a]

[a] Autonomous University of Madrid (Spain)
[b] Institute of Biomedical Research of Lleida (Spain)
[c] University of Lleida (Spain)
[d] Université Paris Ouest La Défense (France)
[e] Institute of Psychology, University of Lausanne (Switzerland)
[f] Psychological Sciences Research Institute, University of Louvain, Louvain-la-Neuve (Belgium

ARTICLE INFO

Keywords:
Personality disorders
DSM-5
PID-5
Factor analysis
Measurement invariance

ABSTRACT

The validity of cross-cultural comparisons of test scores requires that scores have the same meaning across cultures, which is usually tested by checking the invariance of the measurement model across groups. In the last decade, a large number of studies were conducted to verify the equivalence across cultures of the dimensional Alternative Model of Personality Disorders (DSM-5 Section III). These studies have provided information on configural invariance (i.e., the facets that compose the domains are the same) and metric invariance (i.e., facet-domain relationships are equal across groups), but not on the stricter scalar invariance (i.e., the baseline levels of the facets are the same), which is a prerequisite for meaningfully comparing group means. The present study aims to address this gap. The Personality Inventory for DSM-5 (PID-5) was administered to five samples differing on country and language (Belgium, Catalonia, France, Spain, and Switzerland), with a total of 4,380 participants. Configural and metric invariance were supported, denoting that the model structure was stable across samples. Partial scalar invariance was supported, being minimal the influence of non-invariant facets. This allowed cross-cultural mean comparisons. Results are discussed in light of the sample composition and a possible impact of culture on development of psychopathology.

The year 2012 marks an important milestone for the evaluation of psychopathology with the proposal by Krueger et al. (2012) of a publicly available inventory to assess pathological personality (The Personality Inventory for DSM-5; PID-5). These authors operationalized pathological personality as 25 maladaptive personality traits falling within five correlated broad domains, namely Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism. The model was tested for empirical support using exploratory factor analysis techniques. The authors reported pure markers of each domain (e.g., Emotional Lability was related to Negative Affect but not to the other domains), but also that a great number of facets with cross-loadings on other domains (e.g., Restricted Affect was highly related to both the Negative Affect and Detachment domains). In their discussion of the findings, they specifically asked for future studies aiming to replicate that preliminary factor structure in other samples. Following this request, a series of papers were published in different countries addressing this topic. The empirical evidence available to date has been summarized in two recent meta-analysis (Somma et al., 2019; Watters and Bagby, 2018). Both studies supported the idea that the original structure proposed by Krueger et al. had been replicated with guarantees in the different studies, which also denoted consistency of this model across cultures (both meta-analyses comprise more than 20 samples from 12 different countries; Somma et al., 2019; Watters and Bagby, 2018).

However, there is still a question to be addressed: Can the scores be compared between different cultural groups? Measurement invariance of an instrument across a given variable such as culture or sex is a basic

☆Author Note: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this manuscript was supported by Grant PSI2017–85,022-P (Ministerio de Ciencia, Innovación y Universidades, Spain) and the UAM-IIC Chair «Psychometric Models and Applications. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

* Correspondence concerning this article should be addressed to Miguel A. Sorrel, Department of Social Psychology and Methodology, Autonomous University of Madrid, Ciudad Universitaria de Cantoblanco, Madrid 28049, Spain.

E-mail address: miguel.sorrel@uam.es (M.A. Sorrel).

https://doi.org/10.1016/j.psychres.2021.114134
Received 25 September 2020; Received in revised form 16 July 2021; Accepted 19 July 2021
Available online 22 July 2021

requirement for meaningful comparisons of scores on the instrument between the groups formed with that variable. For example, if we compare the means of the observed scores (e.g., sum scores) on the Negative Affect domain for two cultural groups and find a significant difference, ideally we would want to claim that the observed significant difference is due to actual differences on the Negative Affect latent factor (i.e., the construct we intend to measure). The observed score of Negative Affect is only an indicator of that latent score. For this interpretation to be valid (i.e., that the observed difference in Negative Affect is due to a real difference in the latent factor) it is necessary to check at least the following degrees of measurement invariance: 1) the facets that form the Negative Affect domain are the same for both groups (this is referred to as configurational invariance); 2) the factor loadings, which define the degree of relationship between the facets and the Negative Affect domain, are the equal for both groups (this is referred to as metric invariance); and, finally, 3) the baseline level of the facets that form the domain is the same in both groups (this is referred to as scalar invariance). This third level is captured in the intercepts of the indicators (e.g., Emotional Lability) of the latent factor (e.g., Negative Affect) in the factor analysis model. Suppose that for two different cultures A and B the levels of the Emotional Lability facet vary, being much higher in culture A. This would be indicated by the fact that the Emotional Lability intercept is higher in culture A than in culture B. If we subsequently found higher observed scores for culture A in the domain to which Emotional Lability belongs (i.e., Negative Affect) it could be due to the fact that the scores were higher in Emotional Lability. In this example, to interpret the observed differences in Negative Affect as being exclusively due to different scores on the latent factor of Negative Affect would be an incorrect assertion. In order to be able to claim such a thing, it must first be proven that the facets that compose the domain are the same across groups (configurational invariance), that the factor loadings are the equal across groups (metric invariance), and that the intercepts or baseline levels are the equal across groups (scalar invariance). In plain words, measurement invariance implies verifying that the observed score on Negative Affect means the same for the two groups compared. Only in this case is it valid to interpret that the differences found in the observed domain scores are really due to differences in the latent construct that was intended to be measured. On the contrary, the absence of these three degrees of measurement invariance implies that differences on domain observed scores can be total or partially attributed to the differential functioning of the instrument across groups (Chen, 2007; Contractor et al., 2018).

The two previous meta-analyses by Somma et al. (2019) and Watters & Bagby (2018) are very valuable as they address the question of whether the facets that form the domains of the PID-5 are the same across the different samples evaluated. These studies have served to generate a common factorial solution using data from multiple countries, allowing us to determine which facets are assigned to each domain. This corresponds to the level of configurational invariance. To the extent that the factor loadings estimated in the different samples are similar, metric invariance would be supported. However, as mentioned above, it should be noted that the relationship between the observed domain scores and the underlying latent variables is also affected by both the factor loadings and the intercepts. That is, it can only be affirmed that the differences in the observed means are due to the underlying latent factors when there is empirical support for strong forms of measurement invariance. Specifically, it is necessary to check whether the inventory presents scalar invariance (i.e., the intercepts can be assumed to be the same across groups). The interpretation of analyses using scale mean scores, such as *t*-test, ANOVAs, or multilevel analysis, is only valid when this level of invariance is achieved (e.g., Fischer, and Karl, 2019; van de Vijver and Leung, 1997). This psychometric property has been documented in the case of PID-5 for age (Debast et al., 2018), sex (Suzuki et al., 2019), and clinical status (Bach et al., 2018). In all these studies, evidence of scalar measurement invariance has been documented, implying that mean factor scores can be compared across groups. Therefore, lower levels of invariance such as metric invariance (i.e., the stability of the matrix of factor loadings) have also received favorable evidence. In comparison, cross-cultural measurement invariance has received less attention. The aim of the present paper is to fill this gap. In a previous paper, Thimm et al. (2016) analyzed the measurement invariance of the Norwegian version of the PID-5 compared to the United States student sample gathered by Wright et al. (2012). Measurement invariance was overall supported. This article seeks to extend these findings to other cultures. In doing so, we will consider samples from Belgium, Switzerland, France, Spain, and Catalonia. These samples represent different versions of the instrument: the French, the Spanish, and the Catalonian.

Historically, measurement invariance analyses have been carried out from a confirmatory perspective using confirmatory factor analysis (CFA) models. At present, it is possible to use models of exploratory structural equation models (ESEM; Marsh et al., 2009) to perform this analysis. The ESEM approach may be more suitable for situations where the factor structure is complex. It can be understood that CFA is a special case of ESEM where some factor loadings have been set to zero. There is one previous study analyzing PID-5 data using CFA. Specifically, Fossati et al. (2013) evaluated the Italian translation of the PID-5 in two Italian community samples and found support for five-factor model. It is important to remark that a simple structure model did not fit the data well. To get a good fit, they had to free 73 error terms correlations. This may indicate that the unique variances of some facets overlap. Previous research in the area of personality has shown that obtaining good fit indicators is complicated (e.g., Abad et al., 2018). For the specific case of factorial invariance of the PID-5 for variables such as sex or clinical status, reported CFI values have been in the order of 0.90 (e.g., Bach et al., 2018; Suzuki et al., 2019; Thimm et al., 2016). While the value is far from indicating excellent fit following stricter criteria, it has been considered adequate in previous literature and serves as a reference value to assess the fit obtained in the present study.

In summary, the results of the present study will allow us to make recommendations regarding whether it is relevant or not to compare the observed scores for the PID-5 domains across different cultures. Such comparisons are common in areas like personality or psychological disorders (e.g., Medeiros et al., 2017).

1. Method

1.1. Sample and procedure

The present study uses five different samples totaling 4,380 participants. It is important to note that in the text the terms culture and country are used analogously, considering that the extracted samples are defined by a set of shared language, beliefs, values, social norms, and knowledge that is similar within each sample (Ziegenbein et al., 2008). It is common to find such flexibility in the use of the word culture in the literature on instrument adaptations across countries (e.g., Aluja et al., 2020; Thimm et al., 2016). In the case at hand, three samples were collected from 2,532 French-speaking European participants (Roskam et al., 2015): Belgium (1,593), Switzerland (536), and France (403). 25.9% of those participants were men. Age ranged from 18 to 85 years ($M = 27.22$, $SD = 13.28$). The other two samples were from Spain and Catalonia. The first one included 1052 participants (M (age) = 42.01, SD (age) = 15.46, with 44.8% of them being men), and the second one included 796 participants (M (age) = 42.41, SD (age) = 17.52, with 49.6% of them being men).

Specific details about the data collection strategy for the French-speaking European participants can be found at Roskam et al. (2015). Most of the participants were psychology undergraduates from the universities of Poitiers, Liège, Louvain, and Lausanne. Some of the participants belonging to the Belgium sample were recruited by posting announcements on websites, forums, and social networks, and completed the questionnaire online. Participation in the study was voluntary and participants could quit the study at any time they wished. Participants were aware that information collected would be used in a scientific study.

As for the Spain and Catalonia samples, psychology undergraduates from the Autonomous University of Madrid (Castilian language sample) and the University of Lleida (Catalonian language sample) were trained in the theory and instruments of several personality and psychopathological models. Each student had to administer several personality tests and the PID-5 to five people with the following characteristics: one male or female with an age between 18 and 30 years, one male or female between 31 and 40 years, one male or female between 41 and 50 years, one male or female between 51 and 60 years, and one male or female older than 60 years. They were instructed to balance the sex distribution. Participants were informed about the purpose of the research. To increase motivation, personality profiles were returned to all participants indicating a random code that only they knew. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethical commission of the University of Lleida.

1.2. Instruments

1.2.1. PID-5

This inventory is composed of 220 items assessed with a 4-point Likert-type scale ranging from 0 (very false) to 3 (very true). It measures 25 first-order facets that can be grouped into five second-order domains: Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism. The 25 first-order facets are: Emotional lability, Anxiousness, Restricted affectivity, Separation insecurity, Hostility, Perseveration, Submissiveness, Withdrawal, Anhedonia, Depressivity, Intimacy avoidance, Suspiciousness, Manipulativeness, Deceitfulness, Grandiosity, Attention Seeking, Callousness, Irresponsibility, Impulsivity, Rigid perfectionism, Distractibility, Risk taking, Unusual beliefs and experiences, Eccentricity, and Perceptual dysregulation. Facet scores were computed by averaging the items. The French and Spanish versions of the PID-5 have demonstrated appropriate psychometric properties (Gutiérrez et al., 2017; Roskam et al., 2015). The Catalonian version of the PID-5 was translated for this study with specific permission from the American Psychiatric Association. The back translation was sent to K.E. Markon, co-author of the original American PID-5, for review.

1.3. Data analysis

The two meta-analysis studies by Watters and Bagby (2018) and Somma et al. (2019) reported factor structures summarizing the empirical evidence available in past research. Three CFA model structures were evaluated. The *Original* structure included only the paths between facets and domains with loadings higher or equal than 0.30 in Krueger et al. (2012). The other two model structures were developed considering the factor loadings higher or equal than 0.30 reported in Somma et al. and Watters and Bagby. Finally, we also tested an ESEM model. The four models were estimated in each of the five samples. Model fit was assessed using the root mean square error of approximation (RMSEA), and the comparative fit index (CFI). Values greater than 0.90 for CFI and below 0.08 for RMSEA reflect acceptable fit (Hu and Bentler, 1999). When poor fit was observed, we explored the model modification indices (MI) to locate the sources of misfit. Parameters were sequentially released one-by-one. This procedure was evaluated at Jung and Yoon (2016). The authors found that the implementation with the more conservative 99% confidence level led to acceptable Type I error and power rates. To be considered, the MI had to lead to a significant fit improvement in the five samples. This minimizes the risk of capitalization on chance due to characteristics of each sample (e.g., size, heterogeneity). Considering that a large sample size can inflate the size of the MI (Chou and Bentler, 1990), the order of inclusion was determined by the average standardized expected parameter change values (i.e., SEPC; Saris et al., 2009). We sequentially relaxed parameters until fit was acceptable in the five samples.

A multigroup covariance and mean structure analysis was then applied for testing measurement invariance across the five samples. A set of nested models was compared: (A1) configural invariance, in which the same structure was specified for each group, that is, the items that measure the construct are the same; (A2) metric invariance or invariance of factor loadings, which implies that the meaning of the construct is the same in the different groups; (A3) scalar invariance or invariance of intercepts. Achieving this degree of invariance is necessary to compare means as it implies that examinees with the same level in the construct will obtain the same score in the items or facets; and (A4) strict invariance or invariance of facets' unique variance (i.e., the residual error is the same). The criterion of change in CFI (ΔCFI < -0.010) and RMSEA (ΔRMSEA < 0.015) was followed to determine whether the decrement in fit was relevant (Chen, 2007). As will be pointed out below, partial scalar invariance was achieved which allowed the comparison of the latent means. Cohen's d was used to quantify the differences between samples. Values higher than 0.20, 0.50, and 0.80 represented small, medium, and large effects, respectively (Cohen, 1988). The full scalar invariance model did not fit well with the data. Under this situation, Cheng (2008) suggested comparing the parameters of interest (i.e., latent means) obtained with the full and partial invariance models to evaluate the impact of model misfit. If the results are similar, it can be concluded that the lack of invariance has little impact.

All the analyses were conducted using Mplus 8.0 (Muthén and Muthén, 1998–2017) and the `MplusAutomation` R package (Hallquist and Wiley, 2018) using the robust maximum likelihood (MLR) estimator. Codes can be requested from the corresponding author.

2. Results

2.1. Baseline model for measurement invariance testing

The first step was to evaluate whether a common baseline model could be retained for the five samples compared. The different models were estimated in each of the five samples. As can be seen from Table 1, none of the CFA models displayed acceptable fit levels. The ESEM models provided better fit results, with CFI values in the range of 0.847 to 0.923, although model fit was still unacceptable in three of the samples. As the initial model fit was bad, some error-term covariances were released taking the ESEM models as baseline models (see Table 2). MI were generally very large, being 39.45 the average value. The relevance indices were in the range of −0.06 to 0.08. The order of inclusion of the MI was determined by the size of the SEPC index. In total, as indicated in Table 3, 9 additional parameters were required to achieve an acceptable fit in the five samples. These model modifications involved facets referring to related concepts (e.g., Manipulativeness and Deceitfulness; Perseveration and Rigid perfectionism), which provided certain substantive grounds for their inclusion. This model was the reference model in the subsequent analyses.

2.2. Measurement invariance testing

Different increasingly restrictive nested models were evaluated: configural invariance (same structure), metric invariance (same loadings), scalar invariance (same intercepts), and strict invariance (same residual variance). The model fit indicators obtained are reported in Table 4. The configural invariance model presented an acceptable fit which indicated that the factor-facet relationships were identical across samples. When the loadings were fixed to be equal (i.e., metric invariance), the fit remained adequate (ΔCFI $= -0.007$ and ΔRMSEA $= -0.009$). A third step included the restriction of intercepts to be equal across samples. For this model, although the RMSEA remained good, CFI indicated a relevant decrement in terms of fit. A total of 27 intercepts were released in order to achieve a satisfactory model fit (ΔCFI $=$

Table 1
Model Fit.

Sample	Model	χ2	df	p-value	CFI	RMSEA
Belgium	CFA: Krueger et al.	4323.667	252	<0.001	.768	.101
	CFA: Watters and Bagby	3915.127	250	<0.001	.791	.096
	CFA: Somma et al.	4462.288	253	<0.001	.760	.102
	ESEM	2080.525	185	<0.001	.892	.080
Catalonia	CFA: Krueger et al.	2387.136	252	<0.001	.777	.103
	CFA: Watters and Bagby	1980.959	250	<0.001	.819	.093
	CFA: Somma et al.	2268.174	253	<0.001	.790	.100
	ESEM	1090.398	185	<0.001	.905	.078
France	CFA: Krueger et al.	1370.494	252	<0.001	.693	.105
	CFA: Watters and Bagby	1173.703	250	<0.001	.747	.096
	CFA: Somma et al.	1278.769	253	<0.001	.719	.100
	ESEM	743.149	185	<0.001	.847	.087
Spain	CFA: Krueger et al.	2703.446	252	<0.001	.790	.096
	CFA: Watters and Bagby	2160.927	250	<0.001	.836	.085
	CFA: Somma et al.	2517.024	253	<0.001	.806	.092
	ESEM	1086.851	185	<0.001	.923	.068
Switzerland	CFA: Krueger et al.	1749.356	252	<0.001	.736	.105
	CFA: Watters and Bagby	1696.774	253	<0.001	.746	.103
	CFA: Somma et al.	1532.467	250	<0.001	.774	.098
	ESEM	882.034	185	<0.001	.877	.084

−0.009). Thus, partial scalar invariance across samples was supported. Finally, a model fixing the variance of the error terms to be equal across countries did not get a satisfactory fit (ΔCFI = −0.022). The partial scalar invariance is an acceptable condition for comparing means, so no further MI were included. A total of 11 items had invariant intercepts. The maximum number of released intercepts in a facet was 3, and the average across the 25 facets was 1.08. Table 5 includes the parameter estimates for this model.

2.3. Latent factor means comparison

Since at least scalar invariance could be assumed, it was legitimate to compare the means in the PID-5 domains across samples. Effect sizes of the differences between mean latent factors were computed from the model with partial scalar invariance. Belgium was taken as sample of reference. Fig. 1 depicts the factor means and includes the effect size measures. In total, 24 out the 50 comparisons presented an effect size greater than 0.20. The following text lists the different domains according to the effect size measures.

Domains where there were generally no relevant differences. The most similar domain was Psychoticism since no mean difference in this domain reached a small effect size. Similarly, there were generally no differences in Detachment. Only in two comparisons there was an effect size greater than or equal to 0.20: Belgium and France vs. Catalonia, the sample with the lowest mean for this domain.

Domains where there were generally small or medium-sized differences. The average effect size was small for Negative affect and Antagonism (0.33 and 0.20, respectively). For Negative affect, the means for France and Belgium were similar. The other three samples obtained lower and comparable values. Particularly noteworthy is the case of Catalonia, which, as in the case of Detachment, had the lowest mean. As for Antagonism, the differences were smaller. The means for Catalonia and Belgium were similar. France, Switzerland, and Spain obtained lower and comparable means.

The domain with the largest differences. Disinhibition had a medium average effect size (0.46), indicating that in general there was greater variability in the means compared to the other domains. The largest differences occurred for the France-Switzerland, France-Spain, and Belgium-Spain comparisons (d = 0.70, −0.70 and 0.93, respectively). France obtained the highest mean, followed by Belgium. Spain

Table 3
Model Fit (Updated Considering the Local Fit Results).

Sample	Model	χ2	df	p-value	CFI	RMSEA
Belgium	ESEM	2080.53	185	<0.001	.892	.080
	ESEM: MI 8	1307.44	177	<0.001	.936	.063
	ESEM: MI 9	1305.27	176	<0.001	.936	.063
Catalonia	ESEM	1090.40	185	<0.001	.905	.078
	ESEM: MI 8	866.43	177	<0.001	.928	.070
	ESEM: MI 9	841.10	176	<0.001	.931	.069
France	ESEM	743.15	185	<0.001	.847	.087
	ESEM: MI 8	567.65	177	<0.001	.893	.074
	ESEM: MI 9	528.35	176	<0.001	.903	.070
Spain	ESEM	1086.85	185	<0.001	.923	.068
	ESEM: MI 8	846.36	177	<0.001	.943	.060
	ESEM: MI 9	849.28	176	<0.001	.942	.060
Switzerland	ESEM	882.03	185	<0.001	.877	.084
	ESEM: MI 8	621.88	177	<0.001	.922	.068
	ESEM: MI 9	629.92	176	<0.001	.920	.069

Note. MI: Modification index. The number following MI indicates the number of MI that were added.

Table 2
Local Fit Results.

	MI by Sample					SEPC by Sample					
New specification	Belgium	Catalonia	France	Spain	Switzerland	Belgium	Catalonia	France	Spain	Switzerland	SEPC Absolute Row-Mean Order
Y14 WITH Y13	167.86	15.17	8.65	63.84	56.17	.08	.02	.05	.03	.05	1
Y20 WITH Y6	104.88	22.88	43.67	24.00	41.48	.06	.03	.07	.02	.05	2
Y3 WITH Y1	82.73	13.18	19.15	34.27	28.12	−0.06	−0.02	−0.05	−0.04	−0.06	3
Y19 WITH Y1	41.39	10.15	19.44	12.69	20.88	.05	.02	.06	.03	.05	4
Y10 WITH Y9	241.38	59.91	29.57	57.37	57.95	.07	.03	.04	.02	.05	5
Y16 WITH Y15	75.93	28.46	8.83	49.48	31.76	.05	.03	.02	.04	.05	6
Y25 WITH Y19	78.90	20.42	18.39	14.97	8.96	−0.04	−0.02	−0.04	−0.02	−0.02	7
Y23 WITH Y21	15.32	14.79	19.31	8.16	12.59	−0.02	−0.02	−0.04	−0.02	−0.03	8
Y23 WITH Y13	35.04	24.34	7.20	13.66	11.77	.03	.03	.02	.02	.03	9
Y6 WITH Y3	23.36	25.75	12.29	15.22	12.68	.02	.02	.03	.02	.02	10
Y12 WITH Y2	13.55	20.82	7.10	7.30	8.09	.02	.02	.02	.01	.02	11
Y17 WITH Y13	33.35	15.06	11.81	29.09	13.73	−0.02	−0.01	−0.02	−0.02	−0.02	12
Y25 WITH Y14	36.98	38.39	12.98	15.05	21.59	.02	.01	.02	.01	.02	13.5
Y10 WITH Y6	41.79	19.85	9.25	17.70	15.22	−0.02	−0.01	−0.02	−0.01	−0.02	13.5

Note. MI: Modification index. SEPC: Standardized expected parameter change. WITH statements: Correlations between error terms. *Y* is used to denote facets.

Table 4
Measurement Invariance of the PDI-5 Facet Level Structure.

Model	χ2	df	*p*-value	CFI	RMSEA	ΔCFI	ΔRMSEA
(A1) Configural invariance	4189.10	880	<0.001	.932	.066		
(A2) Metric invariance	4895.41	1280	<0.001	.925	.057	−0.007	−0.009
(A3) Scalar invariance	6385.32	1360	<0.001	.896	.065	−0.029	.008
MI: Realising 27 intercepts[a]	5385.408	1333	<0.001	.916	.059	−0.009	.002
(A3) Strict invariance	6558.98	1433	<0.001	.894	.064	−0.022	.005

Note. $\Delta CFI \leq -0.01$ / $\Delta RMSEA \geq 0.015$ denote a relevant difference.
[a] Catalonia: Y7, Y5, Y19, Y2; France: Y15, Y18, Y16, Y17, Y5, Y9, Y7, Y25, Y22, Y2, Y3; Spain: Y19, Y5, Y1, Y7, Y18, Y25; Switzerland: Y2, Y1, Y15, Y6, Y23, Y17.

Table 5
Parameter Estimates for the Partial Invariance Model.

Standardized Factor Loadings						
Facet	Name	Negative Affect	Detachment	Antagonism	Disinhibition	Psychotisim
1	Emotional lability	.615	−0.026[a]	.024[a]	.058	.187
2	Anxiousness	.781	.211	.029	−0.039	.022[a]
3	Restricted affectivity	−0.256	.571	.26	.089	.042
4	Separation insecurity	.695	−0.095	.135	.057	−0.097
5	Hostility	.309	.169	.495	.050[a]	−0.021[a]
6	Perseveration	.469	.267	.065	.091	.244
7	Submissiveness	.37	.180	.038[a]	.157	−0.12
8	Withdrawal	−0.047	.825	.082	−0.017[a]	.071
9	Anhedonia	.166	.610	−0.113	.229	−0.012[a]
10	Depressivity	.352	.372	−0.089	.272	.104
11	Intimacy avoidance	−0.141	.546	.041	.065	.123
12	Suspiciousness	.338	.323	.230	.012[a]	.098
13	Manipulativeness	.002[a]	−0.026	.694	.015[a]	−0.022[a]
14	Deceitfulness	.022[a]	.067	.61	.279	−0.113
15	Grandiosity	−0.023[a]	.106	.667	−0.234	.178
16	Attention seeking	.272	−0.294	.573	.113	−0.023[a]
17	Callousness	−0.185	.300	.564	.203	−0.061
18	Irresponsibility	−0.076	.075	.162	.671	.052
19	Impulsivity	.094	−0.214	.144	.434	.209
20	Rigid perfectionism	.534	.235	.393	−0.637	.214
21	Distractibility	.120	.141	−0.166	.539	.274
22	Risk taking	−0.275	−0.333	.283	.253	.269
23	Unusual Beliefs and experiences	.045[a]	.035[a]	.070	−0.055	.741
24	Eccentricity	.024[a]	.151	.058	.242	.527
25	Perceptual dysregulation	.183	.085	−0.023[a]	.231	.615
Average Covariances (Standard Deviations)						
		Negative Affect	Detachment	Antagonism	Disinhibition	Psychotisim
	Negative Affect		.18 (0.07)	.08 (0.1)	.29 (0.08)	.25 (0.06)
	Detachment			.15 (0.03)	.21 (0.05)	.22 (0.04)
	Antagonism				.45 (0.06)	.36 (0.06)
	Disinhibition					.37 (0.05)

Note. Estimates were standardized taking Belgium as reference.
[a] p-value > 0.05.

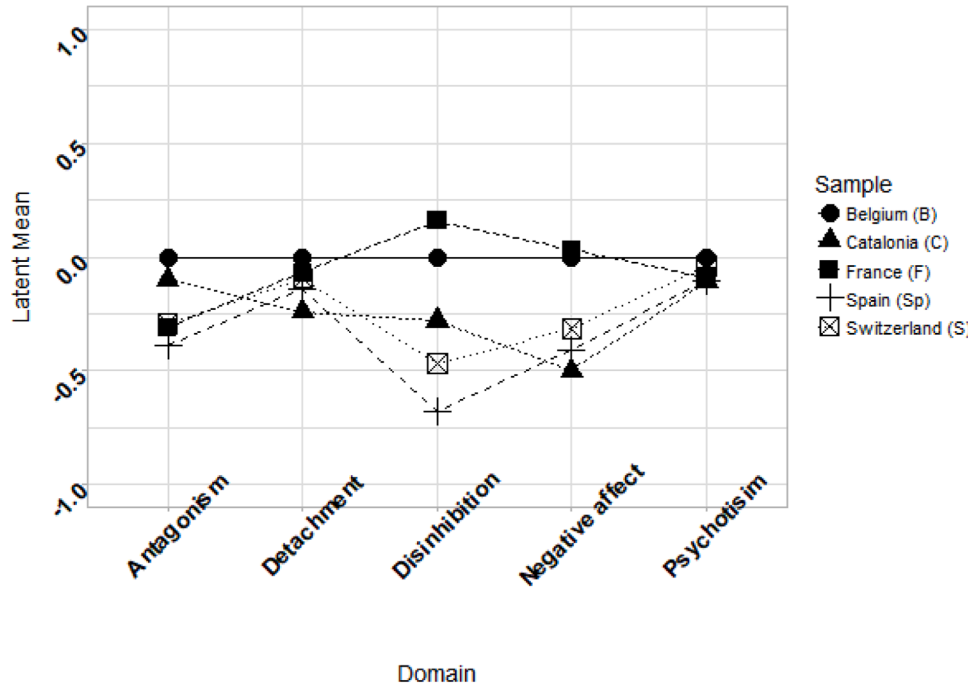
had the lowest mean.

When comparing the differences across all domains, it was found that the most similar samples were Spain and Switzerland (average $\widehat{d} = 0.12$), followed by Belgium and France (average $\widehat{d} = 0.13$). The most dissimilar samples were Belgium and France with respect to Catalonia (average $\widehat{d} = 0.37$ in both cases). The two samples from the same country (Spain and Catalonia) were generally similar, with the notable exceptions of Disinhibition and Antagonism, where Spain had a lower mean.

Finally, in order to assess the consequences of lack of invariance on these mean comparisons, we compared the means in the latent factors imposing the invariance constraint on the non-invariant intercepts (i.e., a full scalar invariance model) with the ones obtained with the partial scalar invariance model. As shown in Fig. 2, these parameters almost had a perfect relationship, which indicates that it may be considered appropriate to make mean comparisons despite not having a full scalar invariance model.

3. Discussion

The present study provides evidence for the measurement invariance of the PID-5 across the cultures examined. This pattern concurs with Thimm et al. (2016)'s results which indicated that the scale was also invariant across United States and Norwegian samples. Since Thimm et al. analyzed university students only, they claimed that a more heterogeneous sample with a larger variety of age, educational level, and socioeconomic status should be analyzed in the future to test the cross-cultural stability of the structure of PID-5. Although university students were also analyzed in the present study, two samples came from the general population. Considering that findings from the present paper are more generalizable, the limitation of the composition of the sample is somewhat surpassed. Note also that no sharp differences in the structure were found between university (French-speaking samples) and community samples (Spanish and Catalonian samples). As in Thimm et al.'s study, some intercepts had to be released. These results imply that if different groups are to be compared, the more appropriate approach would be comparing the latent factor means. However, as the practical influence of the non-invariant items was shown to be minimal, observed scores might be also interpreted with caution. This adds to



Effect size of the differences:

	B-S	B-F	B-C	B-Sp	S-F	S-C	S-Sp	F-C	F-Sp	C-Sp
Negative Affect	0.33	-0.03	0.52	0.45	-0.4	0.21	0.12	0.59	0.54	-0.11
Detachment	0.1	0.07	0.25	0.15	-0.04	0.16	0.05	0.2	0.09	-0.11
Antagonism	0.3	0.32	0.1	0.4	0.02	-0.2	0.1	-0.22	0.08	0.29
Disinhibition	0.48	-0.16	0.29	0.7	-0.7	-0.21	0.22	0.5	0.93	0.44
Psychotisim	0.04	0.09	0.12	0.11	0.06	0.08	0.08	0.03	0.02	-0.01

Fig. 1. Latent Means Comparison.

previous evidence showing that PID-5 domain scores can also be compared by sex and clinical status (Bach et al., 2018; Suzuki et al., 2019).

3.1. *What could explain the differences in intercepts and means?*

Differences in the intercepts and the latent means can be caused by two reasons. The first one would be variations on the sample composition. Note that samples from the community settings (Spanish and Catalonian ones) generally presented the lowest mean differences, and that some of the two largest differences were observed between the French and Belgium (university setting) and both Spanish and Catalonian samples for the Disinhibition domain. Besides, there were large differences in the age distribution. French-speaking samples were younger than the Spanish and Catalonian ones. It is a well-established fact that age has an impact on Disinhibition and related constructs as Sensation seeking (Steinberg et al., 2008), with younger subjects showing higher levels of Disinhibition (Zuckerman, 1994). That would be against other authors like Debast et al. (2018) who state that PID-5 is mostly age neutral. Samples also differed in terms of sex distribution with French-Speaking samples being composed mostly by females, and Spanish and Catalonian samples showing a more balanced distribution. Since women score higher on Neuroticism (Costa et al., 2001) and lower on Disinhibition and Sensation Seeking (Zuckerman, 1994), differences between all French-Speaking samples and Spain and Catalonian can be expected on these two domains. This pattern was observed for the comparisons with Belgium and France, but not with Switzerland. Besides, the previous study by Suzuki et al. (2019) explicitly examined the issue of gender invariance measurement for PID-5, finding that the scale was equivalent across sex. In summary, there is some room to support the hypothesis that the differences in the means may be due to variations in sample composition.

The second explanation of the differences across samples implies the presence of a real impact of the culture in the development of psychopathology. From this standpoint, culture may play a role in determining the exact behavioral and contextual formulation of some maladaptive behaviors and psychopathological manifestations (Terracciano and McCrae, 2006). The fact that differences on factor means were not homogeneous across samples with a similar composition (French-Speaking university samples), and that the samples from the same country (Spain) were generally similar in their latent means, suggests that culture may play a relevant role to explain differences on personality disorders. The five samples came from four different countries (i.e., Belgium, France, Spain, and Switzerland). The most widespread framework for comparing cultures is the six-dimensional classification generated by Hofstede (e.g., Hofstede, 2011). This model postulates six dimensions: power distance, uncertainty avoidance, masculinity vs. femininity, long term vs. short term orientation, and individualism vs. collectivism. Arguably, this last dimension has received the most attention in previous personality literature (e.g., Mulder, 2012; Triandis, 2001). According to Hofstede Insights dataset (Hofstede, 2018), Belgium, France, and Switzerland score similarly in the Individualism-collectivism dimension (75, 71, and 68, respectively), while the score for Spain is markedly lower (51). It has been suggested that a possible consequence of the individualistic culture on personality is to encourage the development of distinct/unique attitudes, self-definition, and striving to attain personal

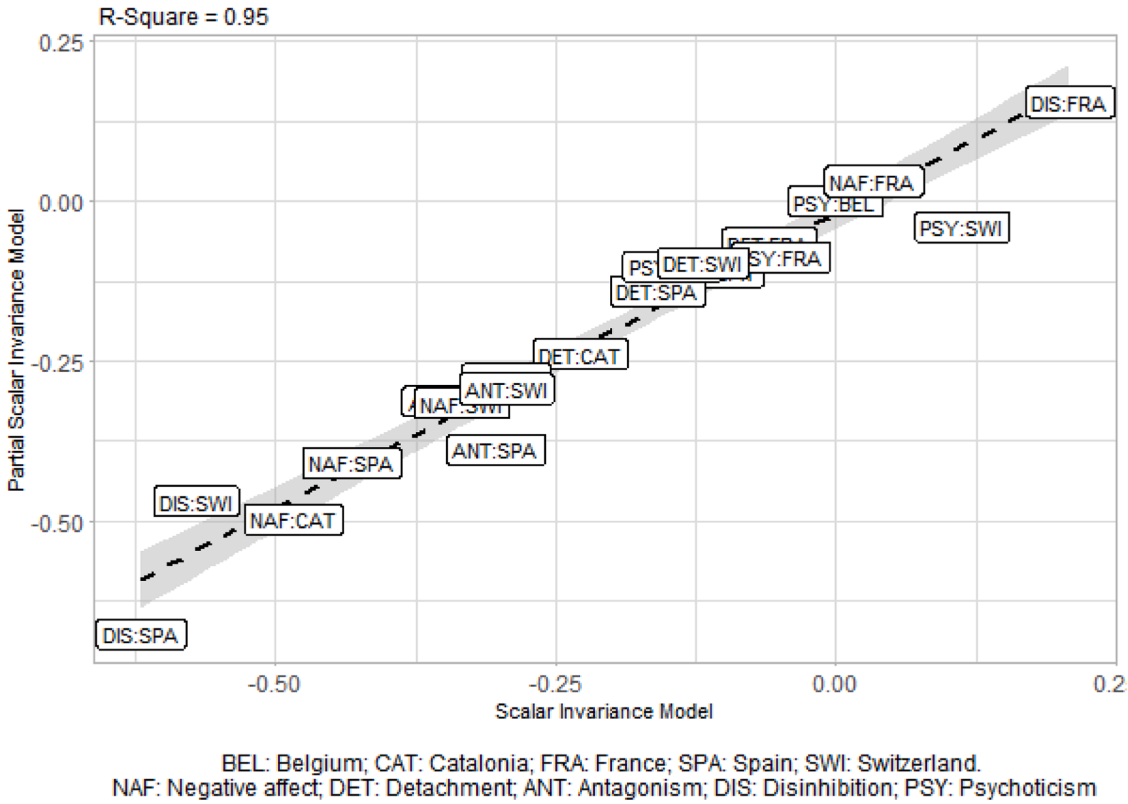


Fig. 2. Scatterplot for the (Full) Scalar and Partial Scalar Invariance Models Scores.

goals (Mulder, 2012). This is somewhat congruent with the fact that France obtained the highest mean in Disinhibition and Spain the lowest. The existence of cohort effects is often cited as an argument for the influence of social and cultural factors on personality. For example, the antisocial behavior profile maintains an upward trend and has doubled in value since World War II (Kessler et al., 1994). However, it should be recalled here that studies linking culture and personality are still scarce, so that the hypotheses put forward should be taken with caution.

3.2. Concluding remarks on the PID-5 factor loading structure

The loading solution was very similar to that reported by the two available meta-analyses (Somma et al., 2019; Watters and Bagby, 2018). It is again evidenced that the factorial structure of PID-5 is complex, with multiple cross loadings. Specifically, 30 out of 100 possible secondary loadings had a factor loading greater than 0.20 in absolute value. Only 8 out of the 25 facets (32%) did not present any secondary loading greater than 0.20 in absolute value (i.e., Emotional lability, Separation insecurity, Submissiveness, Withdrawal, Intimacy avoidance, Manipulativeness, Irresponsibility, and Unusual beliefs and experiences). The most complex facets were Rigid perfectionism and Risk taking. This complex structure might have serious implications for assessment utility (discriminant validity) and theory (Crego et al., 2015; Watters et al., 2018). While some of these secondary loadings can be supported from a theoretical standpoint (e.g., Risk-taking has been related to Negative Affect, Disinhibition and Antisocial behaviour patterns; Aluja et al., 2007), in general, it draws attention to the fact that a revision of the instrument would probably be necessary to reach the most discriminant structure with theoretical meaning.

3.3. Limitations and future directions

The present work is not without limitations. The reported fit values are adequate considering previous literature on the PID-5, but again serve to illustrate that it is difficult to obtain excellent fit values when exploring factor structure in the areas of personality and personality disorders. Authors such as Hopwood and Donnellan (2010) have argued that it is to be expected that it would not be easy to achieve excellent model-data fit, given how easy it is to find items with similar phrasing or other methodology artifacts (e.g., acquiescence). It is expected that modelling these factors will lead to a better fit (Abad et al., 2018). Yet, it is important to remember that this is one of the sources of validity evidence available to support the interpretation and use of the scale scores. The fact that adequate evidence of criterion-referenced validity for the PID-5 scores is reported in prior research also contributes to that purpose (Al-Dajani et al., 2016). As for the analysed samples, all of them came from European Western cultures. In the personality field, Allik et al. (2017) and Aluja et al. (2020) found that similar cultures seem to have similar mean personality profiles. The present study supports this idea also would apply in the context of pathological personality, as most of the effect sizes of differences between countries did not reach a medium effect size, and latent means for Spanish and Catalonian samples were generally similar. It is possible that incorporating non-European Western cultures could change this pattern of cross-cultural stability of the structure, although some other studies suggest a stability beyond western cultures (Rossier & Rigozzi, 2008). Another limitation is that the current study used only a nonclinical sample. However, the available research seems to indicate that this is not a major concern since prior studies found that the results at the domain level obtained in non-clinical samples might be generalized to clinical populations (Bach et al., 2018).

4. Conclusions

In closing, the results presented in this study indicate the presence of a large overlap between facets in different domains. One aim of the dimensional approach was to avoid the great degree of comorbidity reported for categorical dimensional approaches to psychopathological disorders. The present paper suggests information about which facets should be deleted or modified to achieve the desired simple structure, and which show enough discriminant validity across domains to be retained. The findings largely support the stability of factor structure of PID-5 across countries and languages and raises a common structure to be tested in other samples that is congruent with the previous meta-analysis studies. In the present article we addressed a necessary next step that continues the work initiated in these previous meta-analyses, namely, testing scalar invariance (i.e., the invariance of both loadings and intercepts) to support the comparison of scores observed across cultures. One of the most relevant results is that support was obtained for the partial scalar invariance model and that the effect of the non-invariant items was small. This implies that it is legitimate to make comparisons based on mean scores. This result contributes to establish a future dominant position of the dimensional model to replace the current categorical one.

CRediT authorship contribution statement

M.A. Sorrel: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Project administration. **L.F. García:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Project administration. **A. Aluja:** Conceptualization, Investigation, Project administration, Writing – review & editing. **J.P. Rolland:** Conceptualization, Investigation, Project administration. **J. Rossier:** Conceptualization, Investigation, Project administration. **I. Roskam:** Conceptualization, Investigation, Project administration. **F.J. Abad:** Methodology, Software, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abad, F.J., Sorrel, M.A., García, L.F., Aluja, A., 2018. Modeling general, specific, and method variance in personality measures: results for ZKA-PQ and NEO-PI-R. Assessment 25, 959–977. https://doi.org/10.1177/1073191116667547.

Al-Dajani, N., Gralnick, T.M., Bagby, R.M., 2016. A psychometric review of the Personality Inventory for DSM–5 (PID–5): current status and future directions. J Pers Assess 98 (1), 62–81. https://doi.org/10.1080/00223891.2015.1107572.

Allik, J., Church, A.T., Ortiz, F.A., Rossier, J., Hřebíčková, M., De Fruyt, F., McCrae, R.R., 2017. Mean profiles of the NEO personality inventory. J Cross Cult Psychol 48 (3), 402–420. https://doi.org/10.1177/0022022117692100.

Aluja, A., Cuevas, L., García, L.F., García, O., 2007. Zuckerman's personality model predicts MCMI-III personality disorders. Pers Individ Dif 42, 1311–1321. https://doi.org/10.1016/j.paid.2006.10.009.

Aluja, A., Rossier, J., Oumar, B., García, L.F., Bellaj, T., Ostendorf, F., Ruch, W., Wang, W., Kövi, Z., Ścigała, D., Čekrlija, Đ., Stivers, A.W., Di Blas, L., Valdivia, M., Ben Jemaa, S., Atitsogbe, K.A., Hansenne, M., Glicksohn, J., 2020. Multicultural validation of the Zuckerman-Kuhlman-Aluja Personality Questionnaire Shortened Form (ZKA-PQ/SF) across 18 countries. Assessment 27 (4). https://doi.org/10.1177/1073191119831770.

Bach, B., Sellbom, M., Simonsen, E., 2018. Personality inventory for DSM-5 (PID-5) in clinical versus nonclinical individuals: generalizability of psychometric features. Assessment 25 (7), 815–825. https://doi.org/10.1177/1073191117709070.

Chen, F.F., 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal 14 (3), 464–504. https://doi.org/10.1080/10705510701301834.

Chen, F.F., 2008. What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. J Pers Soc Psychol 95 (5), 1005–1018. https://doi.org/10.1037/a0013193.

Chou, C.P., Bentler, P.M., 1990. Model modification in covariance structure modeling: a comparison among likelihood ratio, Lagrange multiplier, and Wald tests. Multivariate Behav Res 25 (1), 115–136. https://doi.org/10.1207/s15327906mbr2501_13.

Cohen, J., 1988. Statistical Power Analysis For the Behavioral Sciences. Academic Press.

Contractor, A.A., Caldas, S.V., Dolan, M., Lagdon, S., Armour, C., 2018. PTSD's factor structure and measurement invariance across subgroups with differing count of trauma types. Psychiatry Res 264, 76–84.

Costa Jr, P.T., Terracciano, A., McCrae, R.R, 2001. Gender differences in personality traits across cultures: robust and surprising findings. J Pers Soc Psychol 81 (2), 322. https://doi.org/10.1037/0022-3514.81.2.322.

Crego, C., Gore, W.L., Rojas, S.L., Widiger, T.A., 2015. The discriminant (and convergent) validity of the Personality Inventory for DSM–5. Personality Disorders: Theory, Research, and Treatment 6 (4), 321–335 https://doi.org/10.1037/per0000118.

Debast, I., Rossi, G., Van Alphen, S.P.J., 2018. Age-neutrality of a brief assessment of the section III alternative model for personality disorders in older adults. Assessment 25 (3), 310–323 http://dx.doi.org/10.1177/1073191118754706.

Fischer, R., Karl, J.A., 2019. A primer to (cross-cultural) multi-group invariance testing possibilities in R. Front Psychol 10, 1507.

Fossati, A., Krueger, R.F., Markon, K.E., Borroni, S., Maffei, C., 2013. Reliability and validity of the Personality Inventory for DSM-5 (PID-5) predicting DSM-IV personality disorders and psychopathy in community-dwelling Italian adults. Assessment 20 (6), 689–708. https://doi.org/10.1177/1073191113504984.

Gutiérrez, F., Aluja, A., Peri, J.M., Calvo, N., Ferrer, M., Baillés, E., Gutiérrez-Zotes, J.A., Gárriz, M., Caseras, X., Markon, K.E., Krueger, R.F., 2017. Psychometric properties of the Spanish PID-5 in a clinical and a community sample. Assessment 24 (3), 326–336. https://doi.org/10.1177/1073191115606518.

Hallquist, M.N., Wiley, J.F., 2018. MplusAutomation: an R package for facilitating large-scale latent variable analyses in Mplus. Structural Equation Modeling: A Multidisciplinary Journal 25 (4), 621–638. https://doi.org/10.1080/10705511.2017.

Hofstede, G., 2011. Dimensionalizing cultures: the Hofstede model in context. Online Readings in Psychology and Culture 2 (1). https://doi.org/10.9707/2307-0919.1014, 2307-0919.

Hofstede, G. (2018). *Hoofstede Insights*. Retrieved from https://www.hofstede-insights.com/country-comparison/belgium,france,spain,switzerland/.

Hopwood, C.J., Donnellan, M.B., 2010. How should the internal structure of personality inventories be evaluated? Personality and Social Psychology Review 14 (3), 332–346. https://doi.org/10.1177/1088868310361240.

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

Jung, E., Yoon, M., 2016. Comparisons of three empirical methods for partial factorial invariance: forward, backward, and factor-ratio tests. Structural Equation Modeling: A Multidisciplinary Journal 23 (4), 567–584. https://doi.org/10.1080/10705511.2015.1138092.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., Kendler, K.S., 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch. Gen. Psychiatry 51, 8–19. https://doi.org/10.1001/archpsyc.1994.03950010008002.

Krueger, R.F., Derringer, J., Markon, K.E., Watson, D., Skodol, A.E., 2012. Initial construction of a maladaptive personality trait model and inventory for DSM-5. Psychol Med 42 (9), 1879–1890. https://doi.org/10.1017/S0033291711002674.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J., Trautwein, U., 2009. Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. Structural Equation Modeling: A Multidisciplinary Journal 16 (3), 439–476. https://doi.org/10.1080/10705510903008220.

Medeiros, G.C., Torres, A.R., Boisseau, C.L., Leppink, E.W., Eisen, J.L., Fontenelle, L.F., Grant, J.E., 2017. A cross-cultural clinical comparison between subjects with obsessive-compulsive disorder from the United States and Brazil. Psychiatry Res 254, 104–111.

Mulder, R.T., 2012. Cultural aspects of personality disorder. In: Widiger, TA (Ed.), Cultural aspects of personality disorder. The Oxford Handbook of Personality Disorders 206–274. https://doi.org/10.1093/oxfordhb/9780199735013.013.0013.

Muthén, L.K., Muthén, B.O., 1998. Mplus User's Guide. –2017, 8th edition. Muthén & Muthén.

Roskam, I., Galdiolo, S., Hansenne, M., Massoudi, K., Rossier, J., Gicquel, L., Rolland, J. P., 2015. The psychometric properties of the French version of the Personality Inventory for DSM-5. PLoS ONE 10 (7). https://doi.org/10.1371/journal.pone.0133413.

Rossier, J., Rigozzi, C., & Personality Across Culture Research Group (16 members) 3, 2008. Personality disorders and the Five-Factor Model among French speakers in Africa and Europe. The Canadian Journal of Psychiatry 53 (8), 534–544. https://doi.org/10.1177/070674370805300808.

Saris, W.E., Satorra, A., Van der Veld, W.M, 2009. Testing structural equation models or detection of misspecifications? Structural Equation Modeling: A Multidisciplinary Journal 16 (4), 561–582. https://doi.org/10.1080/10705510903203433.

Somma, A., Krueger, R.F., Markon, K.E., Fossati, A., 2019. The replicability of the personality inventory for DSM–5 domain scale factor structure in US and non-US samples: a quantitative review of the published literature. Psychol Assess 31 (7), 861–877. https://doi.org/10.1037/pas0000711.

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., Woolard, J., 2008. Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. Dev Psychol 44 (6), 1764–1778. https://doi.org/10.1037/a0012955.

Suzuki, T., South, S.C., Samuel, D.B., Wright, A.G., Yalch, M.M., Hopwood, C.J., Thomas, K.M., 2019. Measurement invariance of the DSM–5 Section III pathological personality trait model across sex. Personality Disorders: Theory, Research, and Treatment 10 (2), 114 http://dx.doi.org/10.1037/per0000291.

Terracciano, A., McCrae, R.R., 2006. Cross-cultural studies of personality traits and their relevance to psychiatry. Epidemiol Psychiatr Sci 15 (3), 176–184. https://doi.org/10.1017/S1121189X00004425.

Thimm, J.C., Jordan, S., Bach, B., 2016. The Personality Inventory for DSM-5 Short Form (PID-5-SF): psychometric properties and association with big five traits and

pathological beliefs in a Norwegian population. BMC Psychol 4 (1), 61. https://doi.org/10.1186/s40359-016-0169-5.
Triandis, H.C., 2001. Individualism-collectivism and personality. J Pers 69 (6), 907–924. https://doi.org/10.1111/1467-6494.696169.
van de Vijver, F.J.R., Leung, K., 1997. Methods and Data Analysis of Comparative Research. SAGE.
Watters, C.A., Bagby, R.M., 2018. A meta-analysis of the five-factor internal structure of the Personality Inventory for DSM–5. Psychol Assess 30 (9), 1255. https://doi.org/10.1037/pas0000605.
Watters, C.A., Bagby, R.M., Sellbom, M., 2018. Meta-analysis to derive an empirically based set of personality facet criteria for the alternative DSM-5 model for personality disorders. Personality Disorders: Theory, Research, and Treatment. https://doi.org/10.1037/per0000307.
Wright, A.G., Thomas, K.M., Hopwood, C.J., Markon, K.E., Pincus, A.L., Krueger, R.F., 2012. The hierarchical structure of DSM-5 pathological personality traits. J Abnorm Psychol 121 (4), 951–957. https://doi.org/10.1037/a0027669.
Ziegenbein, M., Calliess, I.T., Sieberer, M., Machleidt, W., 2008. Personality disorders in a cross-cultural perspective: impact of culture and migration on diagnosis and etiological aspects. Curr Psychiatry Rev 4 (1), 39–47. https://doi.org/10.2174/157340008783743776.
Zuckerman, M., 1994. Behavioral Expressions and Biosocial Bases of Sensation Seeking. Cambridge university press.