Dimensionality, Factorial Invariance, and Cross-Cultural Differential Item Functioning of the Short Dark Tetrad (SD4) in Italian, Romanian, and UK samples

Emanuele Fino[1], Simona A. Popușoi[2], Andrei C. Holman[2], Paolo Iliceto[3], Nadja Heym[1]

Author Note

[1]NTU Psychology, Nottingham Trent University, Nottingham, United Kingdom.

[2]Alexandru Ioan Cuza University of Iasi, Department of Psychology, Iasi, Romania.

[3]Statistics & Psychometrics Ltd, Rome, Italy.

Name, mailing address and email address of author responsible for correspondence about the manuscript:

Dr Emanuele Fino, PhD

NTU Psychology

Nottingham Trent University

NG1 4FQ

Nottingham

United Kingdom

+44(0)115 84 84282

Email: emanuele.fino@ntu.ac.uk

ORCiD: 0000-0002-5095-6014

Dimensionality, Factorial Invariance, and Cross-Cultural Differential Item Functioning of the Short Dark Tetrad (SD4) in Italian, Romanian, and UK Samples

Abstract

The present research tested the dimensionality and cross-cultural differential item functioning of the Short Dark Tetrad (SD4; Paulhus et al., 2021) in samples of university students (total $N = 804$) from Italy ($N = 268$), Romania ($N = 313$), and the United Kingdom ($N = 223$), respectively. Multidimensional Item Response Theory was used to model the data and ordinal logistic regression for differential item functioning analyses. The results showed that a four-factor model was the best fit to the data. However, the model showed metric non-invariance between the three samples, as well as non-negligible differential item functioning for several items in each of the four factors, in line with findings from previous research on cultural differences in antagonistic traits. Implications for theory and research are discussed.

Keywords: Dark Tetrad; personality; assessment; Machiavellianism; narcissism; psychopathy; sadism.

Dimensionality, Factorial Invariance, and Cross-Cultural Differential Item Functioning of the Short Dark Tetrad (SD4) in Italian, Romanian, and UK Samples

Introduction

The 'dark' model of personality was originally defined by Paulhus and Williams (2002) in terms of a set of sub-clinical, stable dispositions to patterns of negative affect, callous, and maladaptive behaviour, commonly referred to as antagonistic or 'dark' traits. Numerous definitions and operationalisations of antagonistic personality exist (e.g., see Kowalski et al., 2021). For example, some authors based the definition of antagonistic traits on the association between an individual's traits and problematic outcomes for self and others (Marcus & Zeigler-Hill, 2015). Other authors pointed out to the co-existence of callousness and interpersonal manipulation, in contrast to honesty-humility (Jones & Figueredo, 2013; Lee et al., 2013). The incremental utility of antagonistic traits was also proposed as a key criterion for a correct classification of them within the personality factor space (Lynam et al., 2006).

One of the most widespread and utilised model of antagonistic personality (see Kowalski et al., 2021, for a review) is Paulhus and Williams' (2002) 'Dark Triad', including Machiavellianism, narcissism, and psychopathy. Machiavellianism can be generally defined in terms of a tendency to premeditation and strategic manipulation of others and of situations to obtain personal gains (Christie & Geis, 1970). Narcissism encompasses a set of feelings of grandiosity, entitlement, risk taking, combined with an overall inflated sense of the self (Kernberg, 1975). Psychopathy refers to a tendency to callousness, impulsive and antisocial behaviour (Hare, 1970). Several measures of antagonistic traits are available. One of the most investigated measures is the 27-item Short Dark Triad (SD3; Jones & Paulhus, 2014). This is a relatively short measure that aimed to overcome the excessive length and theoretical shortcomings of previous antagonistic personality measures, such as the Mach-IV (Christie &

Geis, 1970), the Narcissistic Personality Inventory (Raskin & Hall, 1979), the Self-Report Psychopathy scale (Williams et al., 2007), and the Dirty Dozen (Jonason & Webster, 2010). The SD3 showed excellent psychometric properties, including reliability (Cronbach's alphas values ranging from 0.68 to 0.74) and construct validity (Jones & Paulhus, 2014). The authors used principal axis factor analysis in North American community samples' data (total $N = 1{,}063$), deriving a 27-item scale that confirmed the suitability of their theoretical model, i.e., underlying three positively correlated sub-scales. In fact, results showed that each of the three sub-scales correlated ≥ 0.68 with established measures of the relevant construct, such as the Dirty Dozen (Jonason & Webster, 2010), the Self-Report Psychopathy Scale (SRP-III; Paulhus et al., 2014), and the Mach-IV (Christie & Geis, 1970), supporting the construct validity of the SD3. In addition, the same study found that the SD3 sub-scale circumplex locations matched findings from previous research (e.g., Jones & Paulhus, 2011), in that Machiavellianism and psychopathy presented high dominance/low nurturance, whereas narcissism was associated with high dominance and mid nurturance. However, later studies found relatively high correlations between Machiavellianism and psychopathy, highlighting a risk for redundancy and poor differentiation, challenging the validity of the SD3, with some arguing for the need for further research on the SD3 (Muris et al., 2017).

The Dark Triad later expanded into the 'Dark Tetrad', a model originally proposed in 2009 by Chabrol and colleagues, whereas Paulhus et al. (2021) presented the first comprehensive Dark Tetrad measure, i.e., the Short Dark Tetrad (SD4). Such 4-trait model of antagonistic personality incorporates a distinct factor of sub-clinical sadism, defined as "intrinsic pleasure in hurting others" (p. 208) an individual's drive to harm others, expressed either directly or vicariously (Foulkes, 2019). Paulhus et al.'s (2021) SD4 focussed on vicarious sadism. The sub-scale showed satisfactory construct validity, especially with respect to differentiating sadism and psychopathy. Additionally, they found that sadism was a

significant predictor of sex drive, with those scoring highly in sadism reporting a strong appetite for sex - globally, not just callous sex -, and subsequent research on the SD4 has found a positive correlation between sadism and sociosexual orientation (Blötner, Ziegler, et al., 2022). However, research using profile similarity analysis has indicated a high correlation between the sadism and psychopathy sub-scales (r = .54) in 594 adults, suggesting potential redundancy and concluding that further research is needed to disentangle possible overlaps between the two (Blötner, Ziegler, et al., 2022). Furthermore, a more recent study by Blötner & Mokros found that the sadism and psychopathy sub-scales of the SD4 are at least equally predictive of key correlates of psychopathy and that the psychopathy sub-scale can be considered as a better predictor of key sadism correlates – especially aggression, although the psychopathy sub-scale includes less overt violence-related item content –, carrying substantial implications for the construct validity of the two sub-scales.

The expansion of the Dark Triad into the Dark Tetrad has also contributed to address some of the aforementioned issues related to the construct validity of the SD3, particularly, the need for greater differentiation in the measurement of Machiavellianism and psychopathy. Drawing upon previous research showing that the issue did not seem to affect the Mach-IV items (Christie & Geis, 1970) as much as the SD3 Machiavellianism items (Paulhus & Jones, 2015), the SD4 presented a re-designed and modified version of the Machiavellianism sub-scale and items. Furthermore, in the attempt to control for potential shifting of the factor structure of the scale, after the inclusion of sadism, new items were added to the Machiavellianism, narcissism, and psychopathy sub-scales, as well.

Despite evidence accumulating on the psychometric properties of the SD4, including several studies reporting the adaption of the scale to different cultural and linguistic contexts (e.g., Blötner, Ziegler, et al., 2022), questions on the cross-cultural invariance and Differential Item Functioning (DIF) of the SD4 remain unresolved. In particular, Blötner,

Webster, et al. (2022; see also Blötner, Ziegler, et al., 2022 for an analysis of the nomological network of the SD4) examined the measurement invariance of the SD4 between German and US samples and found configural invariance between the two groups in the overall sample (*Ns* = 594 [German] and 428 [US] after excluding incomplete data from both sub-samples). When restricting the sample to equal age ranges, they found metric invariance (N_s = 170 and 428). Exploratory structural equation modelling identified differences in both samples, compared to the original model, including cross-loadings (≥ 0.3 on more than one factor). Another important finding from the latter study is that some of the intended loadings that had originally been reported by Paulhus et al. (2021) were not replicated, concluding that the original SD4 structure had rather poor fit to the data, and that some of the original items might not be suitable to measure the underlying constructs. Specifically, if on the one hand, two cross-loadings out of 196 were non-negligible, 10 out of the 56 main loadings were negligible (Blötner, Webster, et al., 2022).

Interestingly, the discrepancies found by Blötner, Webster, et al. (2022) were observed across all the sub-scales, however major differences manifested at the level of the sadism sub-scale: Item 6 from the sadism sub-scale non-negligibly cross-loaded onto psychopathy, whereas items 3, 5, 6, and 7 displayed low main loadings. These findings are not surprising, given results from literature showing cross-cultural variation in the manifestation of such traits (Jonason, Foster, Egorova, et al., 2017; Jonason, Foster, Oshio, et al., 2017), and in the same vein, Blötner, Webster, et al. (2022) commented that the observed differences may result from culture-specific norms that regulate individuals' conducts and behaviours as members of specific groups.

Considering the widespread diffusion and popularity of the SD4, investigating its cross-cultural invariance and DIF using an Item Response Theory approach (IRT) is therefore of primary importance to gain a greater understanding of its properties across samples from

different countries, carrying significant implications for theory and research on antagonistic personality. In this vein, a most recent paper by Blötner and Beisemann (2022) evaluated the SD4 using IRT ($N = 594$), showing that the Machiavellianism, narcissism, and psychopathy sub-scales had satisfactory fit and provided a substantial amount of information. Conversely, the sadism items showed poor capacity to discriminate between individuals with similar trait levels and between item difficulty levels, suggesting the need for a revision of the sadism sub-scale. However, to the best of our knowledge, no studies have used Multidimensional Item Response Theory (MIRT) to examine the fit of the hypothesised multidimensional and oblique structure of the SD4, nor DIF.

The question of the invariance of antagonistic traits' measures across cultural and linguistic contexts is key in contemporary personality research, in light of the recent popularity and proliferation of research on such measures across several countries (Dinić & Jevremov, 2021; Jonason, Foster, Egorova, et al., 2017). In this regard, Jonason, Foster, Oshio, et al. (2017) suggested that antagonistic personality traits are sensitive to life history strategies that distinguish and differentiate between cultural contexts. For example, they found country-based differences in patterns of relationships between SD3 scores and individuals' consideration of future consequences (i.e., 'fast' vs. 'low' life history), which they interpreted in terms of country-level variations in individualistic vs. collectivistic orientation, possibly underlying differences in perceived benefits associated with a long-term orientation and disposition to delayed rewards. Similarly, Rogoza et al. (2021) found significant variations in antagonistic traits associated with power distances and hierarchies at the societal level.

The present study included a cross-cultural examination of the SD4, focussed on three European countries, namely Italy, Romania, and the United Kingdom (UK). The locations of these countries in opposite regions of Europe, i.e., Southern, Eastern and Western Europe,

respectively, have been previously characterized in terms of distinct cultural clusters (e.g., Gupta et al., 2002). Accordingly, these countries present substantial differences in their political, institutional, economic and social histories, which have been pinpointed by past research as fostering significant cultural variation (Inglehart & Baker, 2000; Tabellini, 2010).

Cultural differences between European countries have been attested, for instance, by research employing Hofstede et al.'s (2010) model of cultural dimensions (Kaasa et al., 2013; Minkov & Kaasa, 2022), which have identified significant cross-country variations on dimensions which are known to play a role in the manifestation of antagonistic personality, and as a consequence, in its measurement. Among those, individualistic/collectivistic orientation and long-term orientation were found as significantly related to antagonistic traits, with Machiavellianism and psychopathy being associated to ‘faster’ life history strategies and narcissism to ‘slower’ life history strategies, whereas significant country-based differences manifested at the level of individuals’ consideration of future consequences, too (Jonason, Foster, Egorova, et al., 2017).

Specifically, based on the results from the research by Hofstede et al. (2010; see also: Hofstede Insights, 2023; Minkov & Kaasa, 2022), Italian society has been characterised in terms of high levels of individualism associated with a general aversion to authority, along with a tendency to equality and power decentralization. Conversely, Romania has been overall characterised as a collectivistic society, in that individuals tend to privilege long-term social commitments and family/group memberships in their route to fulfillment, with shame being commonly associated with interpersonal offences. Moreover, high levels of power distance characterise the Romanian society, indicative of a general sense of acceptance of the hierarchical order of society. Lastly, the UK locates on the high end of the individualistic continuum, with individuals being motivated by personal goals and being focussed on their

unique contribution to society to minimise societal asymmetries (Hofstede et al., 2010; Hofstede Insights, 2023).

Regarding long-term orientation, Italy has been characterised in terms of a general sense of pragmatism, with situation, context and time considered as key factors in understanding the world and decision-making over one's actions. On the other hand, Romania and the UK present average scores in long-term orientation, midway between valuing adherence to traditional norms while maintaining a pragmatic approach to life (Hofstede et al., 2010; Hofstede Insights, 2023).

Besides cross-cultural variation in the manifestation of antagonistic personality, it is known that cultural norms can also influence how people respond to psychometric scales, for instance, by setting standards of comparison through which respondents evaluate their behavior and subsequently choose their answers to items. Past research has highlighted country-based variations in the prevalence and patterns of mean responses to personality measures. For instance, in their investigation of the personality traits in the Five Factor Model in 22 countries, Kajonius and Mac Giolla (2017) found Romanian participants to score substantially more highly in conscientiousness and lowly in neuroticism in comparison to samples from other European countries, including the UK. Relatedly, Allik et al. (2017) found Italian participants to be lower in neuroticism, extraversion and openness to experience in comparison to Romanian counterparts. Moreover, country-based variations in antagonistic traits have also emerged in past studies, as well as the previously introduced associations between these traits and culture-related dimensions. A recent research by Jonason et al. (2020) found higher levels of narcissism in countries characterised hierarchical cultural systems, whereas the Romanian samples in their investigation scored more highly than the UK samples on all antagonistic traits measured by the Dirty Dozen.

Overall, the substantial differences in geographical location across Europe, associated with cultural differences and potential variations in responding to personality scales that has been suggested by past research, make the three countries of great interest for examining the cross-cultural DIF of the SD4. Consistently, the present research tested the dimensionality, factorial invariance, and DIF of the SD4 in samples from Italy, Romania, and the United Kingdom (UK).

Methods

Participants and procedure

We used convenience sampling to recruit participants. These were 268 students at Italian universities, 313 students at a Romanian university, and 223 university students at a UK university. No multivariate outliers were found by using the factor analytic method proposed by Chalmers and Flora (2015) and the 95th percentile of the $F_{(k+1,\ n-k-1)}$ distribution as cut-off (with k representing the number of observed variables in the model). The final data set comprised a total of 804 observations.

Participation was voluntary and inclusion criteria were: (i) being 18 years old or older; (ii) self-reportedly, being fluent in either Italian, Romanian, or English, depending on the context in which participants had been recruited. Inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. Participants in Romania and the UK took part in the study from September 2021 to June 2022. Participants in Romania were recruited at the end of classroom sessions. Participants in the UK were recruited and compensated with research credits via an online institutional research participation scheme. Regarding the Italian sample, 68 participants were recruited via word of mouth and snowball sampling, from September 2021 to June 2022, whereas the remaining 200 participants were recruited via Prolific (*prolific.co*), an online participants' recruitment platform, in January 2023, being compensated at the rate of £6.03 per hour.

All data files, research, and supplementary materials are available at https://osf.io/rp85n/. The study was not pre-registered.

Measures

Participants completed screening and demographic questions (age, gender) and a psychometric questionnaire including the *Shork Dark Tetrad* (SD4; Paulhus et al., 2021), in Italian (Fino & Iliceto, 2022), Romanian (Bajcsi et al., 2023), or the original English version. Each of the 28 items is rated on a 5-point Likert scale from 1 (i.e., "not at all") to 5 (i.e., "very much").

In developing the Italian and Romanian version of the SD4, we tried to ensure the accuracy of the translations by first having the items translated in each language by two authors of this paper, native speakers of that language. Then, each version was back-translated by a professional native-speaker translator, and the output of this process consisted of a translation that was compared to the original draft. In both cases, no notable discrepancies between the two versions were identified, and therefore, the original draft was used in the study.

Additional information on the psychometric properties of the SD4 are provided in the Introduction and the Results sections of the present manuscript.

Statistical analyses

Multiple-group Multidimensional Item Response Theory (MIRT) was used. All analyses were run in R version 4.1.2 (R Core Team, 2022). A series of Graded Response Models (GRM; Samejima, 1969) were fitted by means of the the *mirt* package (Chalmers, 2012). The GRM is an Item Response Theory model suitable for ordered polytomous items (Samejima, 1969), defining item responses as a function of a single discrimination (slope, α) parameter and category threshold parameters (β). The Stochastic EM algorithm was used for MIRT. This is a two-stage version of the Metropolis-Hastings Robbins-Monro algorithm (Cai,

2010) that uses a stochastically imputed complete-data likelihood estimation (Chalmers, 2012). The multidimensional solution was rotated by using the promax method, aligned with Paulhus et al.'s (2021) procedure. In addition to discrimination parameters, the Item Information Function (IIF) was used to evaluate the performance of the items, indicating the amount of statistical information provided by each item in the measurement of the relevant latent trait. $S\text{-}X^2$ measures (Orlando & Thissen, 2000) and factors' empirical marginal reliability (Chalmers, 2012) were also obtained and evaluated to assess items' fit ($p < 0.001$ as indicative of poor fit) and factors' reliability (acceptable values ≥ 0.70), respectively.

To test the dimensionality of the SD4, Velicer's (1976) minimum average partial method was employed, using inter-item polychoric correlations, unweighted least square estimation, and promax rotation. This method consists of running a series (i.e., k -1) of partial correlation analyses, in which a progressive number of latent dimensions and their relevant items get partialled out to determine a coefficient of off-diagonal average squared correlations. The best number of dimensions to retain corresponds the matrix showing the lowest coefficient value in the series, representing the best trade-off between systematic and unsystematic variance (Revelle, 2020).

To test the factorial invariance of the solution we estimated a series of confirmatory multi-group MIRT/GRM nested models by progressively and cumulatively constraining to equality factor loadings (metric invariance) and (ii) factor loadings and intercepts (scalar invariance) between the samples, respectively. Latent means and variances were fixed (means equal to zero, variances equal to one) for all the groups in the configural and metric model, whereas they were freely estimated for the focal groups and fixed for the reference group in the scalar model (Chalmers, 2012). We evaluated and compared models by using the following fit indices, tests, and cut-off values (Chen, 2007): The Comparative Fit Index (CFI), considering values $\geq .90$ as indicative of satisfactory fit and ≥ 0.010 difference between

models as indicative of non-invariance; the Root Mean Square Error of Approximation (RMSEA), considering values < .06 as indicative of overall satisfactory fit and ≥ 0.015 difference between models as indicative of non-invariance; the Likelihood Ratio Test (LRT), considering a cut-off probability value of 0.001; information criteria such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample size-adjusted version of the BIC (SABIC), with lower values indicating a better fit of the model to the data.

Lastly, we tested for cross-cultural DIF using proportional-odds ordinal logistic regression at the sub-scale level, through the R package *lordif* (Choi et al., 2011). This is a methodology that has been developed for unidimensional measures. However, DIF testing at the sub-scales' level has been suggested to perform as well as with unidimensional models (Chalmers, 2018). We ran three comparisons (focal group vs. reference group): Italy (IT) vs. Romania (RO), RO vs. UK, and IT vs. UK. For each item and a minimum cell count of 5, we compared three nested models: (i) a full model, accounting for the main effect of group membership and the interaction between group membership and trait estimates; (ii) a restricted model, i.e., a model accounting for trait estimates and group membership, but no interaction; (iii) a further restricted model, accounting for no group membership.

Following the recommendations by Choi et al. (2011), we used a general DIF test (2 degrees of freedom) to compare the first and the third of the three nested models, aiming to detect overall (uniform and non-uniform) DIF whilst controlling for Type I error. Then, we ran two DIF tests (1 degree of freedom each) to compare the first model and the second model, then the second and the third model, respectively, to characterise DIF as uniform vs. non-uniform. DIF detection was based on both McFadden's pseudo-R^2 change (Choi et al., 2011) and Likelihood Ratio tests (Alpha = .001). We used the following cut-off values to interpret effect sizes (Jodoin & Gierl, 2001): negligible (< .035), moderate (from .035

to .069), and large (> .070). Additionally, we used Test Characteristic Curves (TCCs) at the level of the SD4 sub-scales to interpret the results, based on Stocking and Lord's (1983) method. This consists of equating the initial and final calibration matrices and identifying the linear constants that minimise the sum of squared differences between the TCCs obtained on non-DIF items over the latent continuum (see Choi et al., 2011, p. 6).

Results

Preliminary data screening

Age and gender distributions by sample are presented in Table 1.

[Table 1: About Here]

Item 6 from the sadism sub-scale showed missing data for response option 5 (i.e., "Strongly Agree") in the Romanian sample, and for this reason, response options 5 and 4 (i.e., "Agree") for that specific item were collapsed. To improve the interpretability of the information reported in the following paragraphs, an overview of the SD4 item contents is provided in Table 2.

[Table 2: About Here]

Following, results of the tests of dimensionality, factorial invariance, and MIRT, and DIF are presented. Detailed descriptive statistics, minimum average partial correlations, fit indices, MIRT parameters, and DIF estimates are provided as supplementary material.

Test of the dimensionality of the SD4

Velicer's (1976) minimum average partial method showed that the four-factor solution had the lowest average squared partial correlation matrix in each of the three samples. The full output of the VSS procedure, including RMSEA and BIC, is provided as supplementary material.

Multiple-group multidimensional item response theory

The multiple-group configural MIRT/GRM showed that the confirmatory four-factor model had satisfactory fit to the data (CFI = 0.918, RMSEA = 0.046 [95% CI = 0.044-0.049]). Regarding the factors' marginal reliability, acceptable (≥ 0.70) to excellent (≥ 0.90) values were found in each of the three samples (IT: Machiavellianism = 0.78, narcissism = 0.90, psychopathy = 0.90, and sadism = 0.87; RO: 0.76, 0.84, 0.80, 0.76; UK: 0.74, 0.84, 0.80, 0.85). All the items showed satisfactory fit (Alpha = .001) across the three samples but item 7 (S-χ^2 = 111.88, df = 67, p < 0.001, RMSEA = 0.050) of the Machiavellianism sub-scale in IT, and item 6 of the Machiavellianism sub-scale in RO (S-$\chi^2 = 97.42_{57}$, S-RMSEA = 0.048, S-p < 0.001). Detailed MIRT/GRM item parameters are reported as supplementary material.

Test of the factorial invariance of SD4

Non-invariance was found at the level of factor loadings (i.e., metric) and intercepts (i.e., scalar), with substantial decrease in all fit indices and LRTs showing low probability levels (p < 0.001) across the nested models, whereas increased values in all the information criteria (i.e., AIC, BIC, SABIC) were observed. Detailed results from the analysis are presented as supplementary materials.

Differential item functioning

We identified several items flagged for DIF (p < 0.001; R^2 ≥ 0.02). Figure 2 displays all the TCCs obtained in each comparison.

[Figure 2: About Here]

IT vs. RO showed overall uniform differential functioning at the level of the Machiavellianism sub-scale, with IT presenting higher expected sub-scale's scores than RO along the trait continuum. The narcissism and psychopathy sub-scales similar TCCs, with minor exceptions for the latter, in which RO showed slightly higher expected values than IT at higher expected values of the trait. The TCC of sadism outlined slightly higher expected values for RO vs. IT at high levels of the trait, but lower expected values at low levels of the

trait. Uniform DIF ($R^2 < 0.02$) was found for items 2 ($R^2 = 0.07$), 6 ($R^2 = 0.09$), and 7 ($R^2 = 0.15$) of the Machiavellianism sub-scale, items 2 ($R^2 = 0.06$), 6 ($R^2 = 0.18$), and 7 ($R^2 = 0.12$) of the narcissism sub-scale, and item 6 ($R^2 = 0.07$) of the psychopathy sub-scale. Negligible ($R^2 = 0.02$) uniform DIF was observed for items 5 of the Machiavellianism sub-scale and item 5 of the psychopathy sub-scale.

Regarding RO vs. UK, TCCs showed higher expected scores for UK along the trait continuum of Machiavellianism. The inspection of the TCCs of narcissism, psychopathy, and sadism showed overall analogous functioning between samples, with some minor exceptions (e.g., lower expected scores for UK vs. RO in psychopathy at higher levels of the trait). Moderate DIF was found for item 3 of the Machiavellianism sub-scale ($R^2 = 0.05$). Uniform but negligible DIF was observed for items 1 ($R^2 = 0.02$) and 2 ($R^2 = 0.03$) of the Machiavellianism sub-scale, item 4 of the narcissism sub-scale ($R^2 = 0.02$), item 3 of the sadism sub-scale ($R^2 = 0.03$) (uniform), and item 6 of the psychopathy sub-scale ($R^2 = 0.02$) (non-uniform).

As for IT vs. UK, the inspection of TCCs across the four SD4 factors showed overall similar item functioning between the two samples, with very minor exceptions. Moderate, uniform DIF was found for item 3 of the psychopathy sub-scale ($R^2 = 0.05$). Negligible, uniform DIF was also observed for items 1 ($R^2 = 0.02$), 3 ($R^2 = 0.03$), and 6 ($R^2 = 0.03$) of the sadism sub-scale. Negligible, non-uniform DIF was finally observed for items 3 of the Machiavellianism sub-scale ($R^2 = 0.03$), 5 and 7 of the narcissism sub-scale ($R^2 = 0.02$), and items 2 and 6 of the sadism sub-scale ($R^2 = 0.02$).

Discussion

The present study examined the dimensionality, factorial invariance, and Differential Item Functioning (DIF) of the Short Dark Tetrad (SD4; Paulhus et al., 2021), a measure of antagonistic personality traits, in three samples of university students from Italy, Romania, and the UK, respectively. The results from the analyses showed that a four-factor model was the best fit to the data, however, the model displayed metric non-invariance and several country-based DIFs.

These results align with findings from previous research. In particular, they confirm the suitability of the four-factor structure of the SD4 in comparison to alternative solutions, that each of the four factors is sufficiently reliable (Paulhus et al., 2021), and that the scale is factorially non-invariant (Blötner, Webster, et al., 2022) between different country-based samples. They also provide a detailed account of the performance of items at the level of their discriminative power and information function, using a Multidimensional Item Response Theory (MIRT) approach. Furthermore, they highlight several DIFs between Italian, Romanian, and UK samples, especially at the level of Machiavellianism, although significant DIFs were found for all the sub-scales. Following, we propose our critical evaluation of these results, based on three main interpretative criteria: (i) cross-cultural differences between the three samples; (ii) potential issues derived from the translation and adaptation of assessments of antagonistic traits; (iii) gender differences.

Regarding cross-cultural differences in antagonistic personality, several studies pointed out to the cultural variation in the expression of such traits. Jonason et al. (2017) argued that antagonistic personality traits are characterised by a focus on immediate life outcomes or a ‘fast’ life history, outlining an evolutionary perspective over antagonistic personality. Accordingly, antagonistic traits might be associated with individuals’ dispositions to seek immediate benefits such as mating and ensure survival (Koehn et al., 2019), so that

individuals scoring highly in antagonistic traits would be more inclined to willingly manipulate others to gain personal benefits, favouring short-term rewards and immediate dominance in spite of long-term cooperation.

Jonason et al. (2017) further interpreted cross-cultural differences in antagonistic traits as possibly underlying variation in the expression of individualistic vs. collectivistic values, with long-term commitment being associated with benefits for both the group and the individual, whereas not delaying rewards would characterise the orientation of those communities who prioritise the individual over the group. In this regard, Hofstede et al.'s (2010) study showed remarkable differences in individualism/collectivism between Italian, Romanian, and UK samples, with Romanian samples being characterised by a tendency to collectivism, thus privileging long-term commitment to community values and norms, vs. Italian and UK samples, which are characterised by substantially higher scores in individualism. In this vein, the large DIF effects that we found in Machiavellianism, particularly for Italian and UK samples compared to the Romanian counterparts, might reflect differences in the manifestations of those traits due to cultural dimensions operating at the country's level. Moreover, as regards narcissism, the higher expected scores in Italian vs. Romanian samples might, as well, reflect the sensitivity of narcissistic personality to country-level social, political, economic, and cultural values, with power dynamics and socio-physical constraints experienced by the populations of the two countries as potentially explaining such differences. Similarly, in the present study, DIFs were also observed in psychopathy and sadism, possibly underlying societal differences in power distance, as well as views over and implications of unpredictable behaviour.

However, interpreting DIF entirely and uniquely in terms of cross-cultural variation in the expression of traits may be speculative, whereas we cannot rule out potential pitfalls deriving from the translation of the SD4 in other languages. In this regard, well-known issue

arise from translating personality scales into new languages, respectively underlying a cross-cultural vs. a psychometric nature: On the one hand, the need for a reliable translation of the test in the a language, and on the other hand, the question as to whether the test is empirically equivalent in the relevant socio-cultural context (Simonsen & Mortensen, 1990). Therefore, undesired fluctuations in meaning and purpose of items may arise due to, for instance, the non-equivalence of linguistic forms as well as the poor relevance of the items in the local cultural context (Paunonen & Ashton, 1998). The rich and multifaceted characterisation of the SD4 items might have determined both, and future research would benefit from considering a review of the original SD4 items to ensure that the four components are sufficiently distinct and adequately represented within a given cultural environment, to match features of the local context that might be associated with variation in the characterisation of antagonistic traits. This would be in line with the conclusions drawn Blötner and Beisemann (2022) in their recent research, focussing, for example, on the need for a “thorough revision of the items of the sadism scale.” (p. 1).

Furthermore, the results from the present study showed large DIFs. In particular, the use of one’s influence 'to get other people on one's side' (i.e., Machiavellianism items), the emphasis on an individual's talent, qualities, and potential to become someone who stands out from the mass (narcissism items), as well as the symbolism associated with and implications of 'getting into dangerous situations' (psychopathy) might all be interpreted through different lenses by individuals partaking to the Italian, Romanian, and UK communities, respectively, potentially and/or partially explaining the observed DIFs, hence requiring caution in the translation of SD4 items across linguistic and cultural contexts. In this vein, future studies translating and adapting the SD4 might benefit from employing exploratory scale development techniques such as cognitive interviewing to inspect the quality of the translation and the adaptability of each item and sub-scale to the local linguistic and cultural

context (Daouk-Öyry & McDowal, 2013). This is an issue that clearly does not restrict to the SD4 but to all antagonistic personality assessments. Consistently, future research should not overlook the possible need for context-specific translations and adaptations of those assessments, as well, accounting for known variations in culture and language and their impact on the expression and assessment of antagonistic traits.

Lastly, the present study did not include gender-based factorial invariance tests and DIF analyses, whereas gender represents another known factor that has been associated with variations in the expression of antagonistic personality. For example, Jonason et al. (2020) found that men scoring highly on psychopathy (i.e., considered as one of the 'fastest' of among antagonistic traits, alongside sadism) were also 'fast' in their life strategies, compared to women. This is in line with the notion that 'faster' traits like psychopathy and sadism carry a promise of adaptive benefit, with potential reproductive gains outperforming costs for men but not women, potentially reflecting differences in adaptive strategies to "compete for limited resources in competitive environments" (Jonason et al., 2020, p. 1263). Additionally, societies with a greater drive to gender equality, such as the UK and Italy in comparison to Romania (European Institute for Gender Equality, 2021), also tend to present variations in forms and expression of control over people's anti-social tendencies, on a continuum characterised by the two poles of subtle interpersonal manipulation and overt coercion, respectively. However, recent studies investigating the factorial invariance of the SD4 have found evidence in support of the gender-based invariance of the scale. For these reasons, and in the attempt to avoid speculation, we recommend future research to test for the factorial invariance and DIF of the SD4 more comprehensively, ideally between samples that account for differences in age, gender, culture, and language. This will allow researchers to shed further light on the determinants of such DIFs and possibly suggest measures to address the factorial non-invariance of the SD4, for example, by means of integrative and/or modified

versions of the items that could target the four antagonistic traits equally well, whilst accounting for local and gender-based variations in their manifestations.

Limitations

Firstly, the relatively small sample size limits the generalisability of the results of the present research, with model complexity and number of parameters influencing the stability of MIRT/GRM parameters estimates (Jiang et al., 2016). Secondly, we tested for cross-cultural but not age or gender invariance of the SD4, a decision based on convenience and the imbalance in gender proportions achieved within the three countries and their respective sample sizes, which might have undermined the results. As already discussed, we consider this a major limitation, since gender and culture are known to interact in the expression of antagonistic personality (Jonason et al., 2020), and for this reason, we recommend future research to attempt to replicate these findings, ideally by targeting samples that include various age ranges and diverse demographic profiles.

Conclusions

The non-factorial invariance and differential item functioning observed in all the four sub-scales of the Short Dark Tetrad (Paulhus et al., 2021) between Italian, Romanian, and UK university students' samples underlie important questions on the cross-cultural expression of antagonistic traits, the translation of antagonistic personality assessments across cultures and languages, and the need for further research to investigate the age- and gender-related invariance of the SD4, aiming to improve the construct validity of the scale and its generalisability to other contexts.

References

Allik, J., Church, A. T., Ortiz, F. A., Rossier, J., Hřebíčková, M., de Fruyt, F., Realo, A., & McCrae, R. R. (2017). Mean Profiles of the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, *48*(3), 402–420. https://doi.org/10.1177/0022022117692100

Fino, E., Popușoi, S. A., Holman, A. C., Iliceto, P., & Heym, N. (2023). *Data, analyses, and supplementary materials for: Dimensionality, Factorial Invariance, and Cross-Cultural Differential Item Functioning of the Short Dark Tetrad (SD4) in Italian, Romanian, and UK samples.* https://osf.io/rp85n/.

Bajcsi, N., Fino, E., Popușoi, S. A., & Holman, A. C. (2023). *Factor Structure and Psychometric Properties of the Romanian version of the Short Dark Tetrad* [Manuscript in preparation].

Blötner, C., & Beisemann, M. (2022). The Dark Triad is dead, long live the Dark Triad: An item-response theoretical examination of the Short Dark Tetrad. *Personality and Individual Differences*, *199*, 111858. https://doi.org/10.1016/j.paid.2022.111858

Blötner, C., & Mokros, A. (2023). The next distinction without a difference: Do psychopathy and sadism scales assess the same construct? *Personality and Individual Differences, 205,* 112102. https://doi.org/10.1016/j.paid.2023.112102

Blötner, C., Webster, G., & Wongsomboon, V. (2022). Measurement Invariance of the Short Dark Tetrad Across Cultures and Genders. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000715

Blötner, C., Ziegler, M., Wehner, C., Back, M. D., & Grosz, M. P. (2022). The Nomological Network of the Short Dark Tetrad Scale (SD4). *European Journal of Psychological Assessment*, *38*(3), 187–197. https://doi.org/10.1027/1015-5759/a000655

Cai, L. (2010). Metropolis-Hastings Robbins-Monro Algorithm for Confirmatory Item Factor Analysis. *Journal of Educational and Behavioral Statistics*, *35*(3), 307–335. https://doi.org/10.3102/1076998609353115

Chabrol, H., Van Leeuwen, N., Rodgers, R., & Séjourné, N. (2009). Contributions of psychopathic, narcissistic, Machiavellian, and sadistic personality traits to juvenile delinquency. *Personality and Individual Differences*, *47*(7), 734–739. https://doi.org/10.1016/j.paid.2009.06.020

Chalmers, P. (2018, February 27). DIF/DTF and multidimensional IRT. *Mirt-Package - Google Groups*. https://groups.google.com/g/mirt-package/c/tRP13GRlQhI/m/mUi3H_rIBgAJ

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, R. P., & Flora, D. B. (2015). faoutlier: An R Package for Detecting Influential Cases in Exploratory and Confirmatory Factor Analysis. *Applied Psychological Measurement*, *39*(7), 573–574. https://doi.org/10.1177/0146621615597894

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

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *Journal of Statistical Software*, *39*(1), Article 1. https://doi.org/10.18637/jss.v039.i08

Christie, C., & Geis, F. L. (1970). *Studies in Machiavellianism*. Academic Press.

Daouk-Öyry, L., & McDowal, A. (2013). Using Cognitive Interviewing for the Semantic Enhancement of Multilingual Versions of Personality Questionnaires. *Journal of Personality Assessment*, *95*(4), 407–416. https://doi.org/10.1080/00223891.2012.735300

Dinić, B. M., & Jevremov, T. (2021). Trends in research related to the Dark Triad: A bibliometric analysis. *Current Psychology*, *40*(7), 3206–3215. https://doi.org/10.1007/s12144-019-00250-9

European Institute for Gender Equality. (2021). *Gender Equality Index (Italy—Romania)*. https://eige.europa.eu/gender-equality-index/2021/country/IT / https://eige.europa.eu/gender-equality-index/2021/country/RO

Fino, E., & Iliceto, P. (2022). *Struttura fattoriale e proprietà psicometriche della versione italiana della Short Dark Tetrad (SD4) [Manuscript in Preparation]*.

Foulkes, L. (2019). Sadism: Review of an elusive construct. *Personality and Individual Differences*, *151*, 109500. https://doi.org/10.1016/j.paid.2019.07.010

Gupta, V., Hanges, P. J., & Dorfman, P. (2002). Cultural clusters: Methodology and findings. *Journal of World Business*, *37*(1), 11–15.

Hare, R. D. (1970). *Psychopathy: Theory research, and implications for society*. Wiley.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations. Software of the mind.* McGraw Hill.

Hofstede Insights. (2023). *Country Comparison: Italy, Romania, United Kingdom*. Hofstede Insights. https://www.hofstede-insights.com/country-comparison/italy,romania,the-uk/

Inglehart, R., & Baker, W. E. (2000). Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, *65*(1), 19–51. https://doi.org/10.2307/2657288

Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00109

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education*, *14*(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2

Jonason, P. K., Foster, J. D., Egorova, M. S., Parshikova, O., Csathó, Á., Oshio, A., & Gouveia, V. V. (2017). The Dark Triad Traits from a Life History Perspective in Six Countries. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01476

Jonason, P. K., Foster, J., Oshio, A., Sitnikova, M., Birkas, B., & Gouveia, V. (2017). Self-construals and the Dark Triad traits in six countries. *Personality and Individual Differences*, *113*, 120–124. https://doi.org/10.1016/j.paid.2017.02.053

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420–432. https://doi.org/10.1037/a0019265

Jonason, P. K., Żemojtel-Piotrowska, M., Piotrowski, J., Sedikides, C., Campbell, W. K., Gebauer, J. E., Maltby, J., Adamovic, M., Adams, B. G., Kadiyono, A. L., Atitsogbe, K. A., Bundhoo, H. Y., Bălțătescu, S., Bilić, S., Brulin, J. G., Chobthamkit, P., Del Carmen Dominguez, A., Dragova-Koleva, S., El-Astal, S., … Yahiiaev, I. (2020). Country-level correlates of the Dark Triad traits in 49 countries. *Journal of Personality*, *88*(6), 1252–1267. https://doi.org/10.1111/jopy.12569

Jones, D. N., & Figueredo, A. J. (2013). The Core of Darkness: Uncovering the Heart of the Dark Triad. *European Journal of Personality*, *27*(6), 521–531. https://doi.org/10.1002/per.1893

Jones, D. N., & Paulhus, D. L. (2011). Differentiating the Dark Triad within the interpersonal circumplex. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions.* (pp. 249–268). John Wiley & Sons, Inc.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Kaasa, A., Vadi, M., & Varblane, U. (2013). European Social Survey as a source of new cultural dimensions estimates for regions. *International Journal of Cross Cultural Management*, *13*(2), 137–157. https://doi.org/10.1177/1470595813485379

Kajonius, P., & Giolla, E. M. (2017). Personality traits across countries: Support for similarities rather than differences. *PLoS ONE*, *12*. https://doi.org/10.1371/journal.pone.0179646

Kernberg, O. (1975). *Borderline Conditions and Pathological Narcissism.* Jason Aronson.

Koehn, M. A., Okan, C., & Jonason, P. K. (2019). A primer on the Dark Triad traits. *Australian Journal of Psychology*, *71*(1), 7–15. https://doi.org/10.1111/ajpy.12198

Kowalski, C. M., Rogoza, R., Saklofske, D. H., & Schermer, J. A. (2021). Dark triads, tetrads, tents, and cores: Why navigate (research) the jungle of dark personality models without a compass (criterion)? *Acta Psychologica*, *221*, 103455. https://doi.org/10.1016/j.actpsy.2021.103455

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, Power, and Money: Prediction from the Dark Triad and Honesty–Humility. *European Journal of Personality*, *27*(2), 169–184. https://doi.org/10.1002/per.1860

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The Perils of Partialling: Cautionary Tales from Aggression and Psychopathy. *Assessment*, *13*(3), 328–341. https://doi.org/10.1177/1073191106290562

Marcus, D. K., & Zeigler-Hill, V. (2015). A Big Tent of Dark Personality Traits. *Social and Personality Psychology Compass*, *9*(8), 434–446. https://doi.org/10.1111/spc3.12185

Minkov, M., & Kaasa, A. (2022). Do dimensions of culture exist objectively? A validation of the revised Minkov-Hofstede model of culture with World Values Survey items and scores for 102 countries. *Journal of International Management*, *28*(4), 100971. https://doi.org/10.1016/j.intman.2022.100971

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The Malevolent Side of Human Nature: A Meta-Analysis and Critical Review of the Literature on the Dark Triad (Narcissism, Machiavellianism, and Psychopathy). *Perspectives on Psychological Science*, *12*(2), 183–204. https://doi.org/10.1177/1745691616666070

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

Paulhus, D. L., Buckels, E. E., Trapnell, P. D., & Jones, D. N. (2021). Screening for Dark Personalities. *European Journal of Psychological Assessment*, *37*(3), 208–222. https://doi.org/10.1027/1015-5759/a000602

Paulhus, D. L., & Jones, D. N. (2015). Chapter 20—Measures of Dark Personalities. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of Personality and Social Psychological Constructs* (pp. 562–594). Academic Press. https://doi.org/10.1016/B978-0-12-386915-9.00020-6

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2014). *Self-Report Psychopathy Scale—4th Edition (SRP-4)*. Multi-Health Systems.

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Paunonen, S. V., & Ashton, M. C. (1998). The Structured Assessment of Personality across Cultures. *Journal of Cross-Cultural Psychology*, *29*(1), 150–170. https://doi.org/10.1177/0022022198291008

R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/.

Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, *45*(2), 590–590. https://doi.org/10.2466/pr0.1979.45.2.590

Revelle, W. R. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. https://cran.r-project.org/package=psych

Rogoza, R., Żemojtel-Piotrowska, M., Jonason, P. K., Piotrowski, J., Campbell, K. W., Gebauer, J. E., Maltby, J., Sedikides, C., Adamovic, M., Adams, B. G., Ang, R. P., Ardi, R., Atitsogbe, K. A., Baltatescu, S., Bilić, S., Bodroža, B., Gruneau Brulin, J., Bundhoo Poonoosamy, H. Y., Chaleeraktrakoon, T., … Włodarczyk, A. (2021). Structure of Dark Triad Dirty Dozen Across Eight World Regions. *Assessment*, *28*(4), 1125–1135. https://doi.org/10.1177/1073191120922611

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*(4, Pt. 2), 100–100.

Simonsen, E., & Mortensen, E. L. (1990). Difficulties in Translation of Personality Scales. *Journal of Personality Disorders*, *4*(3), 290–296. https://doi.org/10.1521/pedi.1990.4.3.290

Stocking, M. L., & Lord, F. M. (1983). Developing a Common Metric in Item Response Theory. *Applied Psychological Measurement*, *7*(2), 201–210. https://doi.org/10.1177/014662168300700208

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic Association*, *8*(4), 677–716. https://doi.org/10.1111/j.1542-4774.2010.tb00537.x

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

Williams, K., Paulhus, D., & Hare, R. (2007). *The four facet structure of psychopathy in non-forensic samples*. Journal of Personality Assessment, *88*, 118–129.

Open Science

Open Science

We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact p values, effect sizes, and 95% confidence or credible intervals.

Open Data

We confirm that there is sufficient information for an independent researcher to reproduce all of the reported results, including codebook if relevant (Fino et al., 2023).

Open Materials

We confirm that there is sufficient information for an independent researcher to reproduce all of the reported methodology (Fino et al., 2023).

Preregistration of Studies and Analysis Plans

This study was not preregistered.

Table 1. Age and gender distributions by country

	Italy (*N* = 268)	Romania (*N* = 313)	United Kingdom (*N* = 223)	Overall (*N* = 804)
Age				
Mean (*SD*)	24.36 (4.39)	20.77 (4.08)	20.52 (2.87)	21.90 (4.27)
Gender				
Female	137 (51.12%)	260 (83.07%)	180 (80.72%)	577 (71.77%)
Male	123 (45.90%)	53 (16.93%)	39 (17.49%)	215 (26.74%)
Non-binary	0 (0.00%)	0 (0.00%)	2 (0.90%)	2 (0.25%)
Prefer not to report	8 (2.99%)	0 (0.00%)	2 (0.90%)	10 (1.24%)

Table 2. Summary of the Short Dark Tetrad (SD4; Paulhus et al., 2021) items' descriptors

Machiavellianism	Narcissism	Psychopathy	Sadism
1. Don't share secrets.	1. Feeling like a natural leader.	1. Being out of control.	1. Fights excites them.
2. Get important people on their side.	2. Talent in persuading people.	2. Fight against authorities.	2. Enjoying violent material.
3. Avoid direct conflict.	3. Group activities as self-centered.	3. Having been in fights.	3. Enjoying seeing other people hurt.
4. Keep a low profile.	4. Feeling special.	4. Immediately dive into things.	4. Enjoying violent sports.
5. Plan to manipulate the situation.	5. Exceptional qualities.	5. Trouble with the law.	5. People deserve to suffer.
6. Use flattery to get people.	6. Expecting to become a star.	6. Likely to get in danger.	6. Being mean on social media.
7. Enjoy successful plans.	7. Likely to show off.	7. Others regret if messing with them.	7. Hurt someone with words.

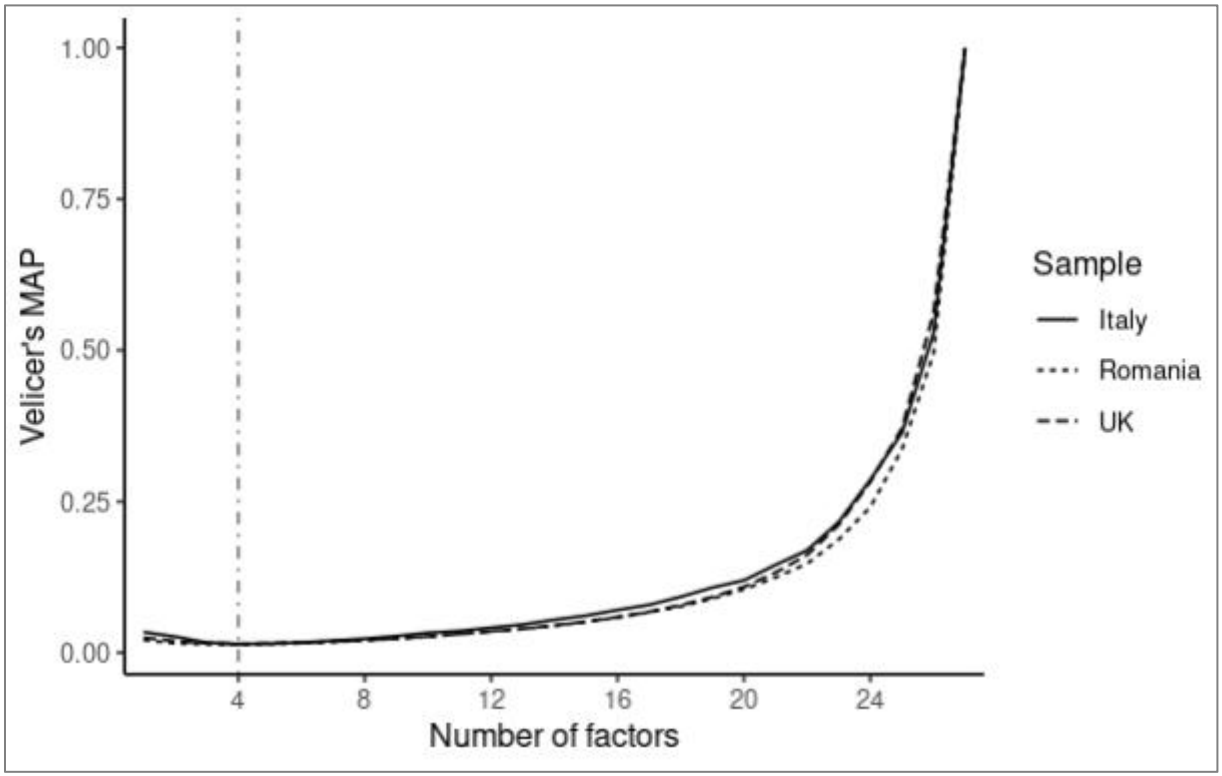
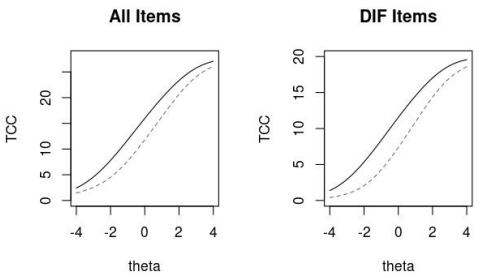
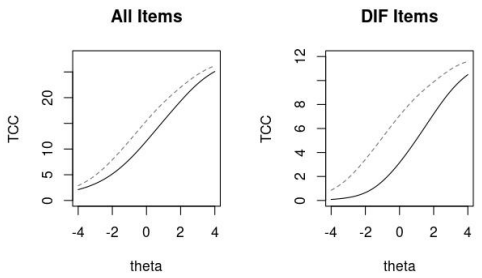
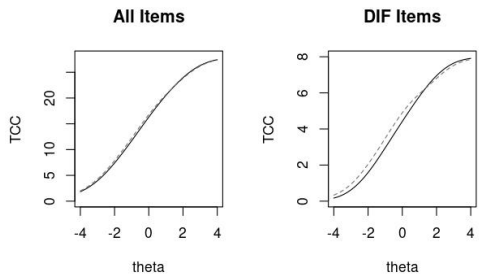
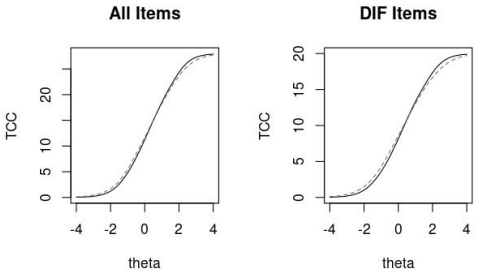
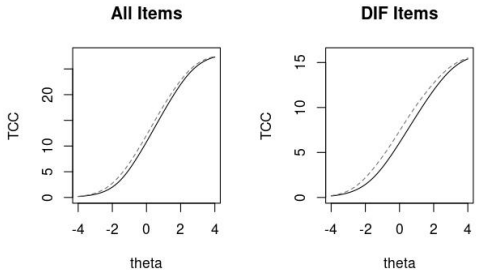
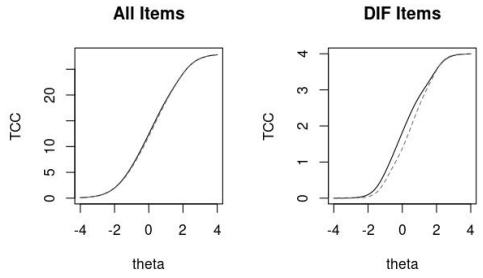
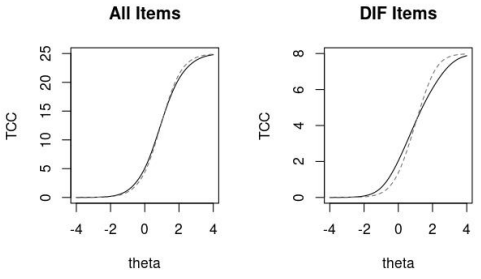
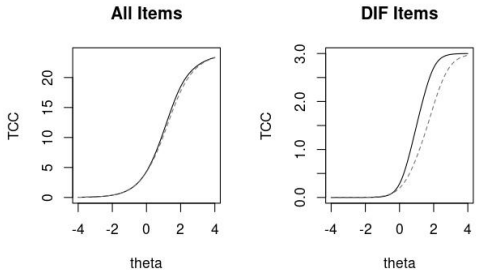
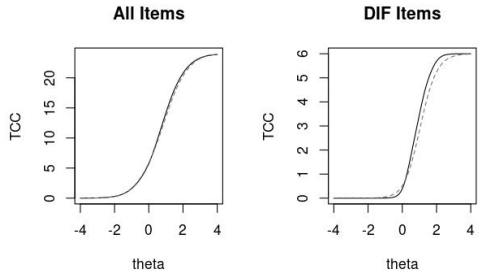


Figure 1. SD4, Velicer's (1976) minimum average partial method for dimensionality testing (Italy: $N = 268$; Romania: $N = 313$; UK: $N = 223$). The vertical line indicates the number of factors that achieved the minimum average, considered as best candidate for retention.

	Italy - Romania	Romania - UK	Italy - UK
Machiavellianism			
Narcissism			
Psychopathy			

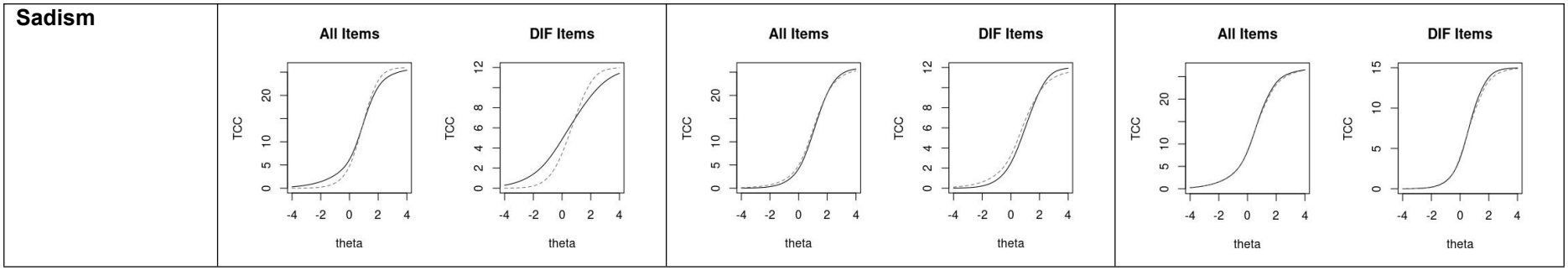


Figure 2. Differential Item Functioning – Test Characteristic Curves, all items vs. flagged items (solid lines represent focal groups).