BMC Psychology

RESEARCH

Open Access

Psychometric validation of the Short Dark Tetrad (SD4): an item-level re-examination with measurement invariance

Burcu Türk[1], Merve Karaburun[2*], Gülçin Şenyuva[3] and Nurcan Hamzaoğlu[4]

Abstract

Psychometric evaluations of the Short Dark Tetrad (SD4; [33]) have been conducted predominantly in Western samples, and the limited evidence from non-WEIRD contexts has yielded inconsistent findings regarding its factor structure. The only prior Turkish adaptation reported that seven items did not function adequately and were removed following exploratory factor analysis, leaving open whether this reflected genuine cultural unsuitability or methodological choices in the analysis of ordinal data. The present study re-examined the SD4 in a community sample of 417 Turkish adults (54.4% female,M_{age} = 30.4, SD = 12.5) using confirmatory factor analysis with the WLSMV estimator and polychoric correlations, the approach recommended for ordinal Likert data. At the item level, only one item (m1, "It is not wise to let people know your secrets") showed a substantive problem (standardized loading = .14, with cross-loadings on Psychopathy and Sadism). The six remaining items removed by Aytaç [1] performed adequately in the present sample (loadings = .53–.74). A 27-item correlated four-factor model excluding m1 showed acceptable fit (CFI = .913, TLI = .904, RMSEA = .069 [.064, .074], SRMR = .070) and outperformed a unidimensional model. Multi-group analyses across gender supported configural, threshold (weak), and threshold-plus-loading (strong) invariance (all $\Delta CFI \leq .005$). Convergent validity was demonstrated through theoretically expected associations with the NPI-13, MPS, LSRP, and Violence Tendency Scale (r = .40–.69). Men scored higher than women on all four traits; effect sizes ranged from small for Narcissism (η^2_p = .010) to moderate–large for Sadism (η^2_p = .115, Cohen's d = 0.72). The findings indicate that the original SD4 structure is largely retained in Turkish samples when ordinal-appropriate estimation is used, with Sadism emerging as the most strongly gender-differentiated trait.

Keywords Dark Tetrad, SD4, Sadism, Measurement invariance, Cross-cultural psychometrics, Ordinal CFA

*Correspondence:
Merve Karaburun
merve.karaburun@btu.edu.tr
[1]Faculty of Arts and Science, Psychology Department, İstanbul Kültür University, Istanbul, Türkiye
[2]Faculty of Humanities and Social Sciences, Psychology Department, Bursa Technical University, Bursa, Türkiye
[3]Faculty of Humanities and Social Sciences, Psychology Department, İstanbul Kent University, Istanbul, Türkiye
[4]Faculty of Arts and Science, Psychology Department, İstanbul Yeni Yüzyıl University, Istanbul, Türkiye

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

In recent years, personality psychology has shown increasing interest in systematically examining tendencies that may harm others. There have been growing scholarly concerns about the psychological mechanisms that underlie manipulative behaviors, lack of empathy, and self-serving behaviors. Within this context, dark personality traits are conceptualized as an umbrella construct encompassing thoughts, emotions, and behaviors that may be harmful to others [14]. These traits are commonly associated with maladjustment, exploitation, egocentrism, and antisocial tendencies, and have been linked to problematic behaviors in both face-to-face and online contexts [19].

The concept of the Dark Triad has long shaped research on dark personality traits. Introduced by Paulhus and Williams [32], this framework includes Machiavellianism, narcissism, and psychopathy. Machiavellianism reflects manipulative tendencies and strategic self-interest,narcissism entails grandiosity and a need for admiration; and psychopathy is characterized by impulsivity, callousness, and lack of remorse [7, 17, 38]. Although these traits are conceptually distinct, they share characteristics such as low empathy, emotional insensitivity, and interpersonal exploitation [14].

The theoretical expansion of this framework led to the inclusion of a fourth trait — sadism — resulting in the Dark Tetrad [31]. Sadism refers to the tendency to derive pleasure from inflicting physical or psychological pain on others [4]. Unlike the other dark traits, sadism is characterized not by instrumental harm but by enjoyment derived directly from the act of harming others. Sadistic tendencies have been found to be particularly evident in online contexts such as internet trolling and digital harassment [5]. More broadly, the Dark Tetrad traits have been associated with interpersonal exploitation, aggression, low self-control, higher levels of risk-taking, and various forms of addiction [20, 29], as well as cyberbullying, intimate partner violence, and online aggression [28, 35].

The growing interest in these traits has increased the need for reliable and valid assessment tools. The Short Dark Tetrad (SD4) was developed by Paulhus et al. [33] as a brief, balanced measure of all four traits with 28 items rated on a 5-point Likert scale. By assessing the four traits within a single instrument, the SD4 provides a comprehensive evaluation of dark personality characteristics. Importantly, the manifestation of dark personality traits and their interpersonal expressions are not independent of cultural context. Social norms, expectations, and interpersonal relationships vary across cultures, potentially influencing how manipulation, lack of empathy, or aggression are expressed. Examining the SD4 across different cultural contexts is therefore essential not only for scale adaptation but also for evaluating the cross-cultural validity of the construct.

Since its development, the SD4 has been examined across a growing number of cultural contexts. Evidence from Portugal, Iran, and Serbia generally supports the original four-factor structure, demonstrating acceptable model fit, adequate reliability, and measurement invariance across demographic groups [2, 10, 34]. In contrast, studies conducted in China reported that several items required revision or replacement to improve cultural appropriateness and psychometric performance, suggesting that some aspects of dark personality assessment may be sensitive to cultural context [26]. Collectively, these findings indicate that although the SD4 shows promising cross-cultural applicability, item-level functioning may vary across societies, highlighting the importance of validating the instrument within each linguistic and cultural setting.

A Turkish adaptation of the SD4 was previously conducted by Aytaç [1] in two samples drawn from a marketing research context ($N = 233$ for EFA and $N = 177$ for CFA). On the basis of exploratory factor analysis, seven items (items 1, 14, 15, 16, 18, 21, and 27 in the original numbering) were excluded because they failed to load adequately on their target factors or cross-loaded onto other dimensions. In addition, the original Machiavellianism and Sadism subscales each split into two subfactors. Aytaç [1] interpreted these results as evidence that the original four-factor structure required substantial modification in a Turkish context.

Taken together, previous cross-cultural findings suggest that differences observed across language versions may reflect not only cultural characteristics but also methodological choices regarding estimation methods and model specification. Further examination of the Turkish SD4 using estimation procedures specifically recommended for ordinal data may help determine whether previously reported structural deviations represent genuine cultural differences or methodological artifacts.

A central question, however, is whether the structural divergence reported by Aytaç [1] reflects genuine cultural unsuitability of these items or methodological factors. SD4 items are rated on a 5-point ordinal scale, and the recommended approach is to use polychoric correlations together with the weighted least squares mean and variance adjusted (WLSMV) estimator [24, 33]. Treating ordinal indicators as continuous in maximum likelihood estimation, factor loadings and fit indices can be biased, particularly with smaller samples and skewed response distributions [39]. Moreover, exploratory factor analysis approaches that remove items on the basis of pattern loadings in a single sample may overfit local sample characteristics, especially when sample sizes fall below standard recommendations for stable factor recovery [27].

These considerations suggest that the items excluded by Aytaç [1] may perform adequately when re-examined using methods for ordinal data in a larger sample.

The present study addressed this question by re-examining the SD4 in a community sample of 417 Turkish adults using item-level confirmatory factor analysis with the WLSMV estimator and polychoric correlations. We tested the original four-factor structure with all 28 items, evaluated the specific items removed by Aytaç [1], and examined whether the original four-factor structure can be supported when methodologically appropriate analyses are conducted. We further evaluated measurement invariance across gender — a prerequisite for the gender comparisons that have featured prominently in the literature on dark traits but were not formally tested in the prior Turkish study. Convergent validity with established measures of psychopathy, narcissism, Machiavellianism, and violence tendency and gender differences in the four traits were also examined.

Method

Participants and procedure

The study protocol was approved by the Social Sciences and Humanities Research Ethics Committee of Istanbul Yeni Yüzyıl University (Approval No. 2025/01–1444; 7 January 2025) and was conducted in accordance with the Declaration of Helsinki and the ethical standards of the institutional research committee. Data collection commenced after ethical approval had been obtained. Participation was voluntary, and informed consent was obtained electronically prior to participation. All study materials, including the informed consent form and questionnaires, were administered online using Google Forms. Google Forms was configured to require a response to every item before allowing progression to the next section; consequently, no item-level missing data occurred within completed questionnaires.

The target population comprised adults aged 18–65 years who were recruited using an online convenience sampling strategy. Data were initially collected from 426 individuals. Nine participants did not complete the survey in full and were excluded prior to analysis, resulting in wholly missing rather than item-level missing data. The final analytic sample resulted in 417 participants, of whom 54.4% were female, and 45.6% were male. More than half of the participants (54.4%) were between 18 and 25 years of age, followed by those aged 35–45 years (15.3%) and 45–55 years (15.1%). Most participants held a university degree (62.8%), while the remainder reported a high school education (23.5%), primary education (7.7%), or postgraduate qualification (6.0%). Most participants were single (67.1%), while the remainder were married (32.9%). As for perceived income, 55.2% reported their income to be approximately equal to their expenses, 27.6% reported lower income than their expenses, and 17.3% reported that their income exceeded their expenses.

Current methodological recommendations indicate that confirmatory factor analysis with ordinal indicators estimated using WLSMV requires a minimum sample size of approximately 200–400 participants for models of moderate complexity [22, 42]. Accordingly, the final sample of 417 participants exceeded these recommendations.

Materials

Data were collected using the Personal Information Form, the Short Dark Tetrad (SD4), the Narcissistic Personality Inventory–13 (NPI-13), the Levenson Self-Report Psychopathy Scale (LSRP), the Machiavellian Personality Scale (MPS), and the Violence Tendency Scale (VTS).

Personal information form

The Personal Information Form was developed by the researchers to collect demographic information. It includes questions about gender, age, education level, income level, and marital status.

Short Dark Tetrad (SD4)

The SD4 [33] is a 28-item self-report measure of Machiavellianism, narcissism, psychopathy, and sadism, with seven items per subscale. Items are rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The original validation study supported the proposed four-factor structure and demonstrated satisfactory internal consistency across the four subscales, with Cronbach's alpha coefficients ranging from 0.75 to 0.81 and McDonald's omega coefficients ranging from 0.76 to 0.81. In the current study, the revised 27-item Turkish version demonstrated good internal consistency, with Cronbach's alpha = 0.90 for the total scale and Cronbach's alpha coefficients of 0.80 for Narcissism, 0.80 for Machiavellianism, 0.78 for Psychopathy, and 0.81 for Sadism.

Narcissistic Personality Inventory–13 (NPI-13)

The NPI-13 [15],Turkish adaptation by [11] is a 13-item forced-choice self-report measure of trait narcissism. The original validation study supported the proposed three-factor structure and demonstrated satisfactory internal consistency (Cronbach's alpha = 0.82). The Turkish adaptation likewise confirmed the original three-factor structure ($\chi^2 = 130$, df = 60, $\chi^2/df = 2.17$, RMSEA = 0.061, CFI = 0.938, TLI = 0.919) and showed good internal consistency for the total scale (Cronbach's alpha = 0.84), with subscale reliability coefficients ranging from 0.61 to 0.78. In the current study, the total score was used ($\alpha = 0.79$).

Levenson Self-report Psychopathy Scale (LSRP)

The LSRP [23],Turkish adaptation by [12] is a 26-item self-report measure assessing primary and secondary psychopathy on a 4-point Likert scale. The original validation study supported the proposed two-factor structure and demonstrated good internal consistency, with Cronbach's alpha coefficients of 0.82 for Primary Psychopathy and 0.63 for Secondary Psychopathy. The Turkish adaptation yielded Cronbach's alpha coefficients of 0.76 for Primary Psychopathy and 0.51 for Secondary Psychopathy, with test–retest reliability coefficients of 0.77 and 0.68, respectively. The total score was used in the present study ($\alpha = 0.75$).

Machiavellian Personality Scale (MPS)

The MPS [9],Turkish adaptation by [41] is a 16-item self-report measure assessing four dimensions of Machiavellianism: amoral manipulation, desire for status, distrust of others, and desire for control. The original validation study supported the proposed four-factor structure. The Turkish adaptation confirmed the original four-factor structure through exploratory and confirmatory factor analyses and demonstrated good internal consistency for the total scale (Cronbach's alpha = 0.88), with subscale reliability coefficients ranging from 0.80 to 0.86. In the current study, the total score was used ($\alpha = 0.84$).

Violence Tendency Scale (VTS)

The VTS [16] is a 20-item self-report measure of aggression and violence tendency rated on a 4-point Likert scale. Higher scores indicate higher levels of aggression and violent tendencies. Previous research demonstrated satisfactory validity and reliability of the scale, with Cronbach's alpha coefficients of 0.78 and 0.87 across two administrations. Split-half reliability analysis also supported the scale's reliability (first-half $\alpha = 0.74$, second-half $\alpha = 0.81$,Spearman–Brown coefficient = 0.86). In the current study, the total score was used ($\alpha = 0.86$).

Data analysis

All analyses were conducted in R version 4.3.3 using the *lavaan* [40] and *semTools* [21] packages. Descriptive statistics, internal consistency (Cronbach's α), and Pearson correlations were computed for all variables. Structural validity was evaluated using confirmatory factor analysis (CFA) with the weighted least squares mean and variance adjusted (WLSMV) estimator and polychoric correlations, treating the 5-point Likert items as ordinal indicators. This estimation method is recommended for ordinal data and is robust to violations of multivariate normality [24, 39].

Analyses were conducted at the item level. We did not use item parcelling, because the central question of the study concerned the adequacy of individual items — including the specific items previously excluded by Aytaç [1]. Although parcelling can improve indicator reliability and yield more stable parameter estimates, it can also obscure item-level misfit, which is the very phenomenon under investigation here [3, 25].

Four competing structural models were compared: (a) a unidimensional model with all 28 items loading on a single latent factor; (b) a four-factor correlated model with all 28 items; (c) a four-factor correlated model with item m1 removed (27 items); and (d) a constrained three-factor model in which Psychopathy and Sadism were merged into a single latent factor, estimated to evaluate the discriminant validity of these two constructs via a scaled chi-square difference test against model (c). Model fit was assessed using the scaled chi-square statistic, the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) with 90% confidence intervals, and the Standardized Root Mean Square Residual (SRMR). Following conventional guidelines, model fit was considered acceptable when CFI and TLI were ≥ 0.90 (preferably ≥ 0.95), RMSEA ≤ 0.08 (preferably ≤ 0.06), and SRMR ≤ 0.08 [18, 22].

Measurement invariance across gender was tested in three nested stages using the procedure of Wu and Estabrook [43] for ordinal indicators: (1) configural invariance (same model structure across groups), (2) threshold invariance (equal thresholds across groups,weak invariance for categorical indicators), and (3) threshold-plus-loading invariance (equal thresholds and factor loadings; strong invariance for categorical indicators). Invariance was supported when ΔCFI did not exceed 0.010, and ΔRMSEA did not exceed 0.015 [6, 37].

Convergent validity was evaluated through Pearson correlations between the SD4 subscale and total scores and the NPI-13, MPS, LSRP, and VTS total scores. Gender differences were examined using one-way analyses of variance; effect sizes were reported as partial eta squared (η^2_p), interpreted as small (0.01), medium (0.06), and large (0.14; [8], and as Cohen's *d*.

Data quality

Following data collection, all completed responses were additionally screened for out-of-range and implausible values. Two minor data adjustments were made before analysis. One participant selected the response option "6" on item m3, which falls outside the SD4 response range (1–5); this value was recoded as missing. The response distribution for item s6 contained an empty cell at the highest category (5) within one gender group; categories 4 and 5 were therefore collapsed into a single category for this item to enable invariance estimation, consistent with standard practice for ordinal CFA [43]. Both adjustments

Table 1 Intercorrelations of SD4 subscales (27-item version) and internal consistency

	Narcissism	Machiavellianism	Psychopathy	Sadism
Total Sample				
Narcissism	1			
Machiavellianism	.39***	1		
Psychopathy	.44***	.35***	1	
Sadism	.37***	.42***	.61***	1
α	.80	.80	.78	.81
α (ordinal)	.84	.84	.85	.88
ω	.82	.82	.81	.85

N = 417. ***p <.001. Machiavellianism reflects the 6-item subscale with m1 removed. Ordinal α calculated from polychoric correlations [44]. *SD4* Short Dark Tetrad

Table 2 Descriptive statistics, reliability, and correlations of SD4 (27-item version) with external measures

Scale	M	SD	α	LSRP	NPI-13	MPS	VTS
SD4 Total (27 items)	67.9	16.4	.90	.530***	.682***	.656***	.693***
Narcissism	19.6	5.52	.80	.326***	.690***	.467***	.438***
Machiavellianism (6 items)	17.0	5.40	.80	.475***	.526***	.593***	.459***
Psychopathy	14.1	5.22	.78	.398***	.435***	.433***	.581***
Sadism	15.2	6.07	.81	.415***	.428***	.502***	.623***

N = 417. ***p <.001. Machiavellianism reflects the 6-item subscale with m1 removed. *LSRP* Levenson Self-Report Psychopathy Scale, *NPI-13* Narcissistic Personality Inventory–13, *MPS* Machiavellian Personality Scale, *VTS* Violence Tendency Scale, *SD4* Short Dark Tetrad

Table 3 Confirmatory factor analysis fit indices for alternative SD4 models

Model	χ^2	df	CFI	TLI	RMSEA [90% CI]	SRMR
1-factor (28 items)	2390.5	350	.719	.696	.118	.117
4-factor correlated (28 items, original)	1061.4	344	.901	.891	.071 [.066,.076]	.075
4-factor correlated (27 items, m1 removed)	943.6	318	.913	.904	.069 [.064,.074]	.070

N = 417. All models estimated with the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator using polychoric correlations for 5-point Likert items. The 27-item model is the recommended specification. *SD4* Short Dark Tetrad

affected at most one cell and did not materially influence model results.

Results

Descriptive statistics and reliability

The Turkish version of the SD4 was administered to 417 adults. Cronbach's α coefficients for the four subscales ranged from 0.78 to 0.81 in the original 28-item subscales and increased to 0.80 for Machiavellianism after removing m1 (see below). Internal consistency for the SD4 total score (27 items) was high (α = 0.90). Because Cronbach's alpha assumes tau-equivalence and can underestimate reliability for ordinal indicators, McDonald's omega coefficients were also computed and ranged from 0.81 to 0.85 across the four subscales, supporting the internal consistency of the scale under a model that allows for varying factor loadings. Ordinal alpha coefficients, calculated from polychoric correlations, were consistently higher than the standard Cronbach's alpha estimates (range = 0.84–0.88), consistent with the expectation that alpha underestimates reliability when applied to ordinal data (Table 1). Observed scores covered nearly the entire theoretical range of the instrument. Table 2 presents descriptive statistics, internal consistency, and correlations with the external measures.

Confirmatory factor analyses

Three CFA models were estimated using WLSMV and polychoric correlations: (a) a unidimensional model, (b) the original four-factor correlated model with all 28 items, and (c) the four-factor correlated model with item m1 removed (27 items). Model fit indices are presented in Table 3.

The unidimensional model showed poor fit (CFI = 0.719, RMSEA = 0.118), indicating that the four traits could not be represented by a single latent dimension. The 28-item four-factor model achieved acceptable fit on RMSEA and SRMR but was below conventional thresholds on CFI and TLI (CFI = 0.901, TLI = 0.891, RMSEA = 0.071 [0.066, 0.076], SRMR = 0.075). Examination of the standardized factor loadings identified one problematic item, m1 ("It is not wise to let people know your secrets"), which showed a very low loading on Machiavellianism (λ = 0.144, p = 0.004). Modification indices indicated that m1 cross-loaded on Psychopathy (MI = 91.5) and Sadism (MI = 85.9) with sizeable negative expected parameter changes, suggesting that this item did not function as a Machiavellianism indicator in the present sample.

When m1 was removed, the resulting 27-item four-factor model fit the data acceptably (CFI = 0.913, TLI = 0.904,

Table 4 Standardized factor loadings and item–total correlations for items removed by Aytaç [1] — item-level four-factor CFA in the present sample (N = 417)

Original item #	Item code	Target factor	Standardized loading	SE	z	*p*	ITC
1	m1	Machiavellianism	.144	.050	2.91	.004	**.23**
14	n7	Narcissism	.603	.037	16.36	<.001	.36
15	p1	Psychopathy	.602	.038	15.73	<.001	.42
16	p2	Psychopathy	.671	.034	19.64	<.001	.55
18	p4	Psychopathy	.533	.040	13.31	<.001	.45
21	p7	Psychopathy	.687	.034	20.16	<.001	.44
27	s6	Sadism	.743	.034	21.79	<.001	.52

Estimates obtained from the 28-item four-factor correlated model with WLSMV and polychoric correlations. Item m1 was the only item to show substantively problematic functioning. *ITC* Item–total correlation within the target subscale (rest-score correlation). *SD4* Short Dark Tetrad

Table 5 Measurement invariance of the 27-item four-factor SD4 across gender (*N* = 417; Female *n* = 227, Male *n* = 190)

Model	χ^2	df	CFI	TLI	RMSEA	SRMR	ΔCFI	ΔRMSEA	LRT p
Configural	1297.3	636	.901	.891	.071	.088	—	—	—
Threshold invariance (weak)	1363.6	689	.899	.897	.069	.088	−.002	−.002	.449
Threshold + Loading invariance (strong)	1355.6	712	.904	.905	.066	.088	+.005	−.003	.783

Invariance testing followed the Wu and Estabrook [43] procedure for ordinal indicators. Invariance was supported when ΔCFI ≤ .010 and ΔRMSEA ≤ .015 [6, 37]. Strong invariance was supported. LRT p = scaled chi-square difference test compared to the preceding model. *SD4* Short Dark Tetrad

RMSEA = 0.069 [0.064, 0.074], SRMR = 0.070). All retained items showed substantial standardized loadings on their target factors (λ = 0.53–0.82, all $p < 0.001$). Latent factor correlations were moderate to large: Machiavellianism–Narcissism $r = 0.51$, Machiavellianism–Psychopathy $r = 0.46$, Machiavellianism–Sadism $r = 0.53$, Narcissism–Psychopathy $r = 0.55$, Narcissism–Sadism $r = 0.45$, and Psychopathy–Sadism $r = 0.77$. The Psychopathy–Sadism correlation was the strongest, consistent with the theoretical proximity of these two constructs, yet remained well below unity, supporting the discriminant validity of the two constructs. To formally evaluate the discriminant validity of Psychopathy and Sadism, a constrained model in which these two factors were collapsed into a single latent factor was compared to the four-factor model via a scaled chi-square difference test. The constrained model showed significantly poorer fit ($\Delta\chi^2 = 75.96$, $\Delta df = 3$, $p < 0.001$), confirming that Psychopathy and Sadism are empirically distinct constructs despite their high latent correlation.

Item-level re-examination of items removed by Aytaç [1]

A central focus of the present study was the seven items removed by Aytaç [1] on the basis of exploratory factor analysis (items 1, 14, 15, 16, 18, 21, and 27 in the original SD4 numbering,m1, n7, p1, p2, p4, p7, and s6 in the present study). The standardized factor loadings and item–total correlations for these items in the present sample are summarized in Table 4.

Six of the seven items exhibited adequate functioning: n7 (λ = 0.60), p1 (λ = 0.60), p2 (λ = 0.67), p4 (λ = 0.53), p7 (λ = 0.69), and s6 (λ = 0.74) (all $p < 0.001$). Item–total correlations for these items ranged from 0.42 to 0.55, within recommended ranges for adequate item discrimination [30]. In contrast, m1 showed poor psychometric performance, with a low standardized loading (λ = 0.14) and a low item–total correlation (0.23), consistent with the cross-loading pattern noted above.

These findings indicate that, with the exception of m1, the items previously excluded from the Turkish SD4 can be retained when ordinal-appropriate estimation is used in an adequately powered sample.

Measurement invariance across gender

Multi-group analyses of the 27-item four-factor model were conducted across gender (female $n = 227$; male $n = 190$). Fit indices for the three nested models are presented in Table 5. The configural model showed acceptable fit (CFI = 0.901, RMSEA = 0.071, SRMR = 0.088). Constraining thresholds to equality across groups (weak invariance) produced a negligible change in fit (ΔCFI = −0.002, ΔRMSEA = −0.002), and additionally constraining factor loadings to equality (strong invariance) likewise produced no meaningful decrement (ΔCFI = +0.005, ΔRMSEA = −0.003). The scaled chi-square difference tests were non-significant at both steps ($p = 0.45$ and $p = 0.78$, respectively). These results support strong measurement invariance, justifying mean-level comparisons of the latent factors across gender.

Subscale intercorrelations and convergent validity

Table 2 presents the correlations between the SD4 subscale and total scores (27-item version) and the external measures. Narcissism showed its strongest correlation with the NPI-13 ($r = 0.69$), Machiavellianism with the MPS ($r = 0.59$), Psychopathy with the VTS ($r = 0.58$) and the LSRP ($r = 0.40$), and Sadism with the VTS ($r = 0.62$). The SD4 total score was moderately to strongly correlated

Table 6 Gender differences on SD4 subscales (27-item version)

Scale	Female M (SD)	Male M (SD)	F	*p*	η^2p	Cohen's d
Narcissism	19.07 (5.11)	20.17 (5.92)	4.17	.042	.010	0.20
Machiavellianism (6 items)	16.24 (5.35)	17.87 (5.32)	9.62	.002	.023	0.31
Psychopathy	13.04 (4.80)	15.28 (5.44)	20.00	<.001	.046	0.44
Sadism	13.33 (4.86)	17.42 (6.52)	53.86	<.001	.115	0.72
SD4 Total (27 items)	61.67 (15.25)	70.77 (17.46)	32.13	<.001	.072	0.56

N = 417 (Female n = 227, Male *n* = 190). Strong measurement invariance was supported (see Table 5), justifying mean-level comparisons. η^2p = partial eta squared; Cohen's d = standardized mean difference (positive values indicate higher male scores). *SD4* = Short Dark Tetrad

with all external criteria ($r = 0.53–0.69$, all $p < 0.001$). The pattern of strongest within-construct correlations and weaker cross-construct correlations supports the convergent validity of each subscale.

Gender differences

Mean-level differences across gender were examined for each subscale (Table 6). Men scored significantly higher than women on all four traits. Effect sizes were small for Narcissism ($\eta^2_p = 0.010$, $d = 0.20$) and Machiavellianism ($\eta^2_p = 0.023$, $d = 0.31$), small-to-moderate for Psychopathy ($\eta^2_p = 0.046$, $d = 0.44$), and moderate-to-large for Sadism ($\eta^2_p = 0.115$, $d = 0.72$). The gender effect on Sadism was approximately five times the effect on Machiavellianism and over ten times the effect on Narcissism, identifying Sadism as the most strongly gender-differentiated trait within the SD4 framework.

Discussion

The present study re-examined the psychometric properties of the Short Dark Tetrad (SD4) in a Turkish community sample of 417 adults, with three specific aims: to test the original four-factor structure at the item level using ordinal-appropriate estimation, to revisit the items removed in the prior Turkish adaptation by Aytaç [1], and to evaluate measurement invariance across gender. The findings provide a coherent and largely positive picture of the SD4 in Turkish, while also identifying one item that does not function adequately and warrants attention in future use of the scale.

Item-level replication of the four-factor structure

At the item level, the four-factor correlated model showed acceptable fit and clearly outperformed a unidimensional alternative. Of the seven items previously excluded by Aytaç [1], six performed adequately in the present sample, with standardized loadings ranging from 0.53 to 0.74 and item–total correlations from 0.42 to 0.55. The remaining item, m1 ("It is not wise to let people know your secrets"), displayed a substantively poor loading on its target factor (0.14) and cross-loaded on Psychopathy and Sadism, indicating that it does not function as a Machiavellianism indicator in this sample. With m1 removed, the resulting 27-item four-factor model achieved acceptable fit on all indices and preserved the originally proposed structure.

This pattern suggests that the structural differences reported by Aytaç [1] were driven primarily by methodological factors rather than by genuine cultural unsuitability of the excluded items. Three considerations are relevant here. First, treating ordinal Likert items as continuous in maximum-likelihood-based estimation can bias factor loadings and inflate the apparent prevalence of weakly loading items [39]. Second, the EFA-based exclusion procedure used by Aytaç [1] operated on combined samples that were smaller than current recommendations for stable factor recovery [27] and may have capitalized on sample-specific variance. Third, the present sample was over twice the size of Aytaç's CFA sample, yielding more stable parameter estimates. Taken together, these considerations suggest that the original SD4 structure is largely retained in Turkish when appropriate analytic choices are made.

These findings are broadly consistent with previous cross-cultural validation studies. Similar to the Portuguese [34], Persian [2], and Serbian [10] adaptations, the present findings support the original multidimensional conceptualization of the SD4. At the same time, evidence from the Chinese adaptation suggests that some items may require modification to better reflect culture-specific expressions of dark personality traits [26]. Taken together, these findings indicate that the overall four-factor framework of the SD4 appears to be robust across diverse cultural settings. However, specific items may function differently depending on cultural and linguistic context.

The primary exception in the present study was item m1. Across all analyses, this item consistently demonstrated weak performance, including a very low loading on Machiavellianism, substantial cross-loadings on Psychopathy and Sadism, and a low item–total correlation. One possible explanation is that the item reflects cautiousness or self-protection rather than strategic interpersonal manipulation within the Turkish cultural context. Interestingly, similar concerns regarding culture-specific item functioning have also been reported in other non-Western adaptations of the SD4. Although the Persian validation retained the original item set, the

Chinese adaptation required modification and replacement of several items to achieve satisfactory psychometric performance, suggesting that certain indicators of dark personality may not transfer equivalently across cultural contexts [2, 26]. Accordingly, the present findings further support the importance of evaluating item functioning at the cultural level rather than assuming complete cross-cultural equivalence. We recommend that future Turkish applications of the SD4 either drop m1 or interpret Machiavellianism scores that include m1 with caution. A 27-item version of the Turkish SD4 with m1 omitted yielded improved subscale reliability for Machiavellianism ($\alpha = 0.80$ vs. 0.78 with m1 included) and is the version we recommend going forward.

Measurement invariance and gender differences

A second contribution of the present study is the formal evaluation of measurement invariance across gender. Configural, threshold, and threshold-plus-loading invariance were all supported ($\Delta CFI \leq 0.005$, $\Delta RMSEA \leq 0.003$), indicating that the SD4 measures the same constructs in the same metric in Turkish-speaking men and women. This is a prerequisite for substantive interpretation of mean-level gender differences and was not tested in the prior Turkish adaptation.

Mean-level analyses revealed that men scored higher than women on all four traits. The most notable finding was the magnitude of the gender effect on Sadism ($\eta^2_p = 0.115$, Cohen's $d = 0.72$), which substantially exceeded the gender effects on the other three traits and falls within the moderate-to-large range. This pattern is consistent with prior reports that gender differences in dark traits are most pronounced for sadistic and antagonistic tendencies [5, 13, 36] and adds cross-cultural evidence from a non-Western sample. By contrast, gender differences on Narcissism and Machiavellianism were small ($\eta^2_p = 0.010$ and 0.023), suggesting that whatever cultural forces shape these traits in Turkey operate roughly equally on men and women.

Convergent validity and subscale intercorrelations

Correlations with established external measures supported the convergent validity of the SD4 subscales. As expected, each subscale demonstrated its strongest association with the conceptually corresponding criterion measure: Narcissism with the NPI-13, Machiavellianism with the MPS, and Sadism with the Violence Tendency Scale. Psychopathy was most strongly correlated with the VTS (0.58) and moderately with the LSRP (0.40), consistent with the conceptual emphasis of the SD4 Psychopathy items on impulsivity and risk-taking rather than the affective–interpersonal features captured more extensively by the LSRP. Subscale intercorrelations followed the expected pattern, with the strongest association between Psychopathy and Sadism (latent $r = 0.77$), reflecting their shared antagonistic core, while remaining well below unity and formally confirmed as discriminable via a scaled chi-square difference test consistent with their conceptualization as distinct dimensions.

Limitations and future directions

Several limitations should be acknowledged. First, participants were recruited using an online convenience sampling strategy and consisted predominantly of younger, university-educated adults, limiting the generalizability of the findings to older, less-educated, clinical, and forensic populations. Online data collection may also introduce selection bias by excluding individuals with limited access to or familiarity with digital technologies, while preventing researchers from directly observing participants or controlling for potential distractions during survey completion. Second, self-report assessment of socially undesirable traits may be influenced by impression management or social desirability. Third, although measurement invariance was supported across gender, invariance across age and educational groups remains to be examined. Future research should therefore evaluate the SD4 in more diverse community samples as well as in clinical and forensic populations to further establish the generalizability of the Turkish version across different segments of society.

Conclusion

The Turkish version of the Short Dark Tetrad, with item m1 removed, appears to be a psychometrically sound instrument for assessing dark personality traits within a multidimensional framework. The findings extend existing evidence on the SD4 by demonstrating that the original four-factor structure can be largely preserved in a non-WEIRD sample when ordinal-appropriate estimation is applied. Furthermore, the study establishes strong measurement invariance across gender and identifies Sadism as the most strongly gender-differentiated trait in the Turkish cultural context.

Acknowledgements

The authors would like to thank the Project Support Office at Bursa Technical University for their assistance in proofreading the English language of this manuscript.

Authors' contributions

Merve Karaburun (Corresponding Author): Methodology, Statistical Analysis, Writing—Reviewing and Editing. Burcu Türk (Co-author): Conceptualization, Methodology, Writing—Original draft Preparation. Gülçin Şenyuva (Co-author): Writing—Reviewing and Editing. Nurcan Hamzaoğlu (Co-author): Writing—Reviewing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability
The datasets generated for this study are available, at reasonable request, to the corresponding author.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Social Sciences and Humanities Research Ethics Committee of Istanbul Yeni Yüzyıl University (Approval No: 2025/01–1444, Date: 07 January 2025) and was conducted in accordance with the ethical standards of the institutional research committee and the Declaration of Helsinki.

Consent for publication
Not applicable. The manuscript does not contain any individual person's identifying data, images, or videos.

Competing interests
The authors declare no competing interests.

Received: 11 May 2026 / Accepted: 20 July 2026
Published online: 28 July 2026

References

1. Aytaç MB. Karanlık Dörtlü (The Short Dark Tetrad–SD4) ölçeğinin Türkçeye uyarlanması ve karanlık dörtlünün pazarlama araştırmalarındaki yerinin tartışılması. İşletme Ekonomi ve Yönetim Araştırmaları Dergisi. 2022;1:132–51. https://doi.org/10.33416/baybem.1041678.
2. Bagajan KQ, Ziegler M, Soleimani M, Paulhus DL, Soleimani ZA, Kordbagheri M, Alavinejad L, Amiri H, Asl VY, Hoseini S, Bagajan HQ. Validation of the Short Dark Tetrad (SD4) in Persian: Assessment of its structure and nomological network. J Individ Differences. 2024. https://doi.org/10.1027/1614-0001/a000417. Advance online publication.
3. Bandalos DL, Finney SJ. Item parceling issues in structural equation modeling. In: Marcoulides GA, Schumacker RE, editors. New developments and techniques in structural equation modeling. Mahwah (NJ): Lawrence Erlbaum Associates; 2001. p. 269–96. https://doi.org/10.4324/9781410601858-15.
4. Buckels EE, Jones DN, Paulhus DL. Behavioral confirmation of everyday sadism. Psychol Sci. 2013;24(11):2201–9. https://doi.org/10.1177/0956797613490749.
5. Buckels EE, Trapnell PD, Paulhus DL. Trolls just want to have fun. Personality Individ Differ. 2014;67:97–102. https://doi.org/10.1016/j.paid.2014.01.016.
6. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model. 2007;14(3):464–504. https://doi.org/10.1080/10705510701301834.
7. Christie R, Geis FL. Studies in Machiavellianism. Academic Press; 1970.
8. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
9. Dahling JJ, Whitaker BG, Levy PE. The development and validation of a new Machiavellianism scale. J Manag. 2009;35(2):219–57. https://doi.org/10.1177/0149206308318618.
10. Dinić BM, Buckels EE, Kovačević N. The Short Dark Tetrad (SD4): Measurement invariance across Serbian and Canadian samples and further validation. Eur J Psychol Assess. 2024**.** Advance online publication. https://doi.org/10.1027/1015-5759/a000864.
11. Doğan U, Çolak TS. Narsisistik Kişilik Envanteri-13'ün (NPI-13) Türkçe formunun psikometrik özellikleri. Anadolu Psikiyatri Dergisi. 2020;21(2):167–74.
12. Engeler A. Psikopati ve antisosyal kişilik bozukluğu [PhD thesis]. Istanbul: Istanbul University, Institute of Forensic Medicine; 2005.
13. Foulkes L, McMillan D, Bell AM. The short sadistic impulse scale: reliability, validity and the role of sadism in trolling behaviors. Personal Individ Differ. 2014;65:31–5.
14. Furnham A, Richards SC, Paulhus DL. The dark triad of personality: a 10-year review. Soc Personal Psychol Compass. 2013;7(3):199–216. https://doi.org/10.1111/spc3.12018.
15. Gentile B, Miller JD, Hoffman BJ, Reidy DE, Zeichner A, Campbell WK. A test of two brief measures of grandiose narcissism: the Narcissistic Personality Inventory–13 and the Narcissistic Personality Inventory–16. Psychol Assess. 2013;25(4):1120–36. https://doi.org/10.1037/a0033192.
16. Göka E, Bayat B, Türkçapar MH. Şiddet eğilimi ölçeğinin geliştirilmesi Düşünen Adam. 1995;8:11–5.
17. Hare RD. Manual for the Revised Psychopathy Checklist. 2nd ed. Toronto (ON): Multi-Health Systems; 2003.
18. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis. Struct Equ Model. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
19. Jonason PK, Webster GD. The dirty dozen: a concise measure of the dark triad. Psychol Assess. 2010;22(2):420–32. https://doi.org/10.1037/a0019265.
20. Jones DN, Paulhus DL. Differentiating the Dark Triad within the interpersonal circumplex. In: Horowitz LM, Strack S, editors. Handbook of interpersonal psychology: theory, research, assessment, and therapeutic interventions. Hoboken (NJ): Wiley; 2011. p. 249–67. https://doi.org/10.1002/9781118001868.ch15.
21. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. semTools: Useful tools for structural equation modeling (R package version 0.5–6). 2022. https://CRAN.R-project.org/package=semTools.
22. Kline RB. Principles and practice of structural equation modeling. 4th ed. New York: Guilford Press; 2016.
23. Levenson MR, Kiehl KA, Fitzpatrick CM. Assessing psychopathic attributes in a noninstitutionalized population. J Pers Soc Psychol. 1995;68(1):151–8.
24. Li C-H. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods. 2016;48(3):936–49. https://doi.org/10.3758/s13428-015-0619-7.
25. Little TD, Cunningham WA, Shahar G, Widaman KF. To parcel or not to parcel: exploring the question, weighing the merits. Struct Equ Model. 2002;9(2):151–73.
26. Liu Y, Zhou B, Ouyang Y, Yang B, Xie Q. Development and validation of Chinese form Short Dark Tetrad (C-SD4). Heliyon. 2023;9(1):e12929. https://doi.org/10.1016/j.heliyon.2023.e12929.
27. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychol Methods. 1999;4(1):84–99.
28. Moor L, Anderson JR. A systematic literature review of the relationship between dark personality traits and antisocial online behaviours. Pers Individ Dif. 2019;144:40–55. https://doi.org/10.1016/j.paid.2019.02.027.
29. Muris P, Merckelbach H, Otgaar H, Meijer E. The malevolent side of human nature: a meta-analysis and critical review of the literature on the Dark Triad (narcissism, Machiavellianism, and psychopathy). Perspect Psychol Sci. 2017;12(2):183–204. https://doi.org/10.1177/1745691616666070.
30. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
31. Paulhus DL. Toward a taxonomy of dark personalities. Curr Dir Psychol Sci. 2014;23(6):421–6. https://doi.org/10.1177/0963721414547737.
32. Paulhus DL, Williams KM. The dark triad of personality: narcissism, Machiavellianism, and psychopathy. J Res Pers. 2002;36(6):556–63. https://doi.org/10.1016/S0092-6566(02)00505-6.
33. Paulhus DL, Buckels EE, Trapnell PD, Jones DN. Screening for dark personalities: the Short Dark Tetrad (SD4). Eur J Psychol Assess. 2020;37(3):208–22. https://doi.org/10.1027/1015-5759/a000602.
34. Pechorro P, Karandikar S, Carvalho B, DeLisi M, Jones DN. Screening for dark personalities in Portugal: intra- and interpersonal correlates, reliability and invariance of the Short Dark Tetrad Portuguese version. Deviant Behav. 2023;44(4):551–66. https://doi.org/10.1080/01639625.2022.2071655.
35. Pineda D, Galán M, Martínez-Martínez A, Campagne DM, Piqueras JA. Same personality, new ways to abuse: how Dark Tetrad personalities are connected with cyber intimate partner violence. J Interpers Violence. 2022;37(13-14):NP11860–NP11884. https://doi.org/10.1177/08862605211015265.
36. Plouffe RA, Wilson CA, Saklofske DH. Examining the relationships between childhood exposure to intimate partner violence, the Dark Tetrad of personality, and violence perpetration in adulthood. J Interpers Violence. 2022;37(5-6):NP3449–NP3473. https://doi.org/10.1177/0886260520948517.
37. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev. 2016;41:71–90. https://doi.org/10.1016/j.dr.2016.06.004.
38. Raskin R, Terry H. A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. J Pers Soc Psychol. 1988;54(5):890–902. https://doi.org/10.1037/0022-3514.54.5.890.
39. Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol Methods. 2012;17(3):354–73. https://doi.org/10.1037/a0029315.

40. Rosseel Y. Lavaan: an R package for structural equation modeling. J Stat Softw. 2012;48(2):1–36. https://doi.org/10.18637/jss.v048.i02.
41. Ülbeği İD. Makyavelizm Kişilik Ölçeği'nin Türkçe formunun geçerlik ve güvenirlik çalışması. Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi. 2016;25(2):89–104.
42. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas. 2013;73(6):913–34. https://doi.org/10.1177/0013164413495237.
43. Wu H, Estabrook R. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. Psychometrika. 2016;81(4):1014–45. https://doi.org/10.1007/s11336-016-9506-0.
44. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales. J Mod Appl Stat Methods. 2007;6(1):21–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.