The Dark Five - statistical anomaly or persistent phenomena: An exploratory investigation into the factor structure of the SD4 in multiple samples

A. Dulović [a] ⓘ, C. Smyth [b,*] ⓘ, D. Waldeck [c] ⓘ

[a] *University of Montenegro, Montenegro*
[b] *University of Staffordshire, UK*
[c] *Coventry University, UK*

ABSTRACT

Recent analysis of the dark tetrad has tentatively proposed that a more appropriate representation of dark traits, consisting of 22 items loading onto a 5-factor solution could resolve acknowledged methodological issues. Replication was needed to determine if this new model (D5-SD4) is a persistent phenomenon or a statistical artefact related to the original data. This study compared model fit and measurement invariance of original SD4 and D5-SD4 in 12 samples from 11 different countries ($N = 7045$). The D5-SD4 exhibited better fit whether 28 items were used for the SD4, or only the shared 22. Measurement invariance and alignment optimization showed that D5-SD4 could be used to compare results between groups, which was not the case for the original SD4. The results demonstrate the reliability of the D5-SD4 structure in comparison to the SD4. Suggestions for the future improvements of the core dark traits model are given.

1. Introduction

Research into so-called dark traits was popularised by Paulhus and Williams (2002), who introduced the Dark Triad, this conceptualisation consisted of three distinct, yet overlapping constructs: Machiavellianism (manipulative, instrumental behaviour; Christie & Geis, 1970), Narcissism (an inflated sense of self-importance; Jones & Paulhus, 2014), and Psychopathy (impulsive callous disinhibition; Jones & Paulhus, 2014). The model was later expanded to include Sadism (the enjoyment of others' suffering, Paulhus et al., 2021).

In relation to traditional personality models that often overlooked these traits, the Dark Tetrad offers significant value in understanding and predicting socially aversive behaviours, such as aggression (Paulhus et al., 2021), stalking (Branković et al., 2023), and criminal behaviour (Međedović et al., 2024). As such, the development and use of robust measures for assessing Dark Tetrad traits is crucial.

To be aligned with previous research on SD4, the word "dark" is used to describe constitutive traits. However, we recognise that using this term can be stigmatizing, inaccurate and raise problematic associations

(Chester et al., 2025; Stanton et al., 2025; Kay & Arrow, 2022). It is also worth noting these traits have been associated with positive outcomes (Aghababaei & Błachnio, 2015; Papageorgiou et al., 2019). We agree with Chester et al. (2025) that new terms for these traits should be explored.

1.1. Measurement crisis

Recent research has raised concerns regarding the measurement of dark traits (Crawford et al., 2025; Kay & Arrow, 2022; Kowalski et al., 2021), with Kowalski et al. highlighting the Jingle and Jangle fallacies. The Jingle fallacy, which Thorndike (1913) attributed to Professor Aikins, occurs when different constructs share the same label, whereas the Jangle fallacy (Kelley, 1927) occurs when essentially the same constructs are given different labels. Since Dark tetrad traits overlap to some extent, these fallacies are somewhat inevitable.

Paulhus claimed that these concerns had been mitigated in the Short Dark Tetrad Scale[1] (SD4; Paulhus et al., 2021); however, Blötner et al. (2022) still found significant overlap between psychopathy and sadism.

---

* Corresponding author at: Department of Psychology, School of Health, Education Policing and Science, University of Staffordshire, Stoke-on-Trent, Staffordshire, UK.

E-mail address: chris.smyth@staffs.ac.uk (C. Smyth).

[1] From this point onwards, the term SD4 will to refer to the scale and not the factor structure which will be referred to as the four-factor model (4FM).

They also found the overlap between other instruments designed to measure individual traits of the SD4, which points to a wider measurement issue Crawford et al. (2025) raised further concerns about the use of fit indices based on item parcelling since that practice can mask multidimensionality (Bandalos, 2002). Should the original non-parcelled fit statistics have been reported for the 4FM, the scale would have failed to meet established cutoffs (Hu & Bentler, 1999).

Previous research on measurement invariance of the SD4 scale showed mixed results. Dinić et al. (2024) found partial scalar invariance between Serbian and Canadian samples, but did not use scaled fit indices. Blötner et al. (2023) reported variations in measurement invariance between US and German samples, while Fino et al. (2025) reported metric non-invariance between Italian, Romanian and UK samples, along with differential item functioning across all four factors. This suggests that factor structure and measurement invariance represent important challenges for the SD4. Consequently, Crawford et al. (2025) utilized a data-driven approach (exploratory SEM analysis) to critically examine the SD4 factor structure and omitted several redundant items. They reported that the 4FM exhibited poor fit, beyond what could be attributed to cross loadings. Further analysis of the scales led them to conclude that two items should be omitted from the narcissism scale and that it should be reconceptualised as grandiose exhibitionism. Sadism was divided into violent voyeurism (the enjoyment of watching violence) and indirect sadism (the enjoyment of inflicting pain without physical intervention), differentiating *spectators* and *actors* when concerning the enjoyment of others' suffering. The label "indirect" comes from the fact that Paulhus et al. (2021) intentionally removed items that referred to direct physical sadism, due to their overlap with psychopathy. In addition, items such as "I know how to hurt someone with words alone" do not necessarily imply direct verbal conflict, since hurting with words can occur in cyberspace, where no direct contact is needed. They also found sadism and psychopathy exhibited substantial overlap with the Machiavellianism factor, and posited that a refined factor better expressed the data. Overall, the authors found that a five-factor model (referred to as the *Dark Five of the SD4*; D5-SD4), consisting of coalition building, grandiose exhibitionism, psychopathy, indirect sadism and violent voyeurism exhibited better fit and that the two factors derived from sadism differentially predicted attachment.

1.2. Current study

The findings of Crawford et al. (2025) represent significant advancement in the understanding of the psychometric structure of the SD4. Moreover, their findings suggested that all factors of the SD4 except psychopathy should be represented differently, with sadism split into two distinct components. A robust measurement model is essential to accurately capture dark traits, warranting further examination of the D5-SD4. For example the current conceptualisation of sadism may be too broad and consequently mask important and potentially directionally differential manifestations of subsumed constructs. Dimensional representations, especially those generated using data-driven techniques may reflect statistical anomalies. To test if the D5-SD4 was simply a statistical anomaly within one dataset (Crawford et al., 2025; Waldeck et al., 2024), it is important to examine the replicability of this structure in other datasets. The newly developed 22-item model of the SD4 (referred to as the D5-SD4) will be examined and compared to the original 28-item, four-factor SD4 model in terms of model fit, factor loadings, and measurement invariance, using secondary datasets from several countries. It was hypothesised that the D5-SD4 would exhibit better fit than the 4FM in relation to fit statistics, and that its factor loadings would more appropriately reflect dimensional representations. Additionally, we sought to demonstrate invariance across various datasets. Establishing measurement invariance will contribute to the understanding of the usability of the SD4 across cultures and different groups to determine whether the D5-SD4 retains important characteristics of model fit in various populations.

2. Methods

2.1. Participants

Secondary datasets were utilized that focused on validation studies of the SD4. Item-level response data were available for all samples and were re-analysed in the present study. Participants responded to SD4 items that were translated into their native language.

In total, there were 7157 participants, from 11 different countries. Datasets were gained from Peru (Zegarra-López et al., 2024; $N = 617$), Serbia (Dinić et al., 2024; $N = 488$), Canada (Dinić et al., 2024; $N = 739$), the Netherlands (Rassin et al., 2024; $N = 751$), Slovakia (Teličák et al., 2024; $N = 1902$), Turkiye (Aytac, 2024; $N = 337$), Poland (Gajda et al., 2022; $N = 251$), 2 samples from Germany (Blötner et al., 2022; Blötner & Grüning, 2023; $N = 598$ & $N = 500$ respectively), the UK (Fino et al., 2025; $N = 223$), Romania (Fino et al., 2025; $N = 313$), and the USA (Webster & Wongsomboon, 2020; $N = 451$). All data used were publicly available, except for the Peruvian data, which was requested from the authors directly.

In the Romanian sample, due to non-convergence, outliers which were more than 3 times the interquartile range were identified and removed, resulting in 13 removals and a final sample of 300.

2.2. Materials

2.2.1. The Short Dark Tetrad Scale (SD4; Paulhus et al., 2021)

This scale consists of 28 items, measured on 5-point Likert scales. Originally, Paulhus et al. (2021) proposed that it consisted of 4 constructs, namely Machiavellianism, Narcissism, Psychopathy and Sadism, however Crawford et al. (2025) proposed an alternative factor structure, consisting of 5 factors, namely, Coalition Building, Grandiose Exhibitionism, Indirect Sadism, Violent Voyeurism and Psychopathy. Table 1 shows the items and how they map onto the 2 proposed factor structures.

2.3. Analysis

Both 4FM and D5-SD4 were tested for measurement invariance, as suggested by Van Der Schoot et al. (2012). This involved testing each sample individually, then combining the samples and sequentially testing more constrained models using Confirmatory Factor Analysis (CFI).

Notably, any direct comparison between the models should take into account that item removal may contribute to improved fit, as might differences in factor structure. To mitigate this, interpretations of fit are made via established cutoffs, and direct comparison must be done cautiously. Note that in supplementary material, extra analyses are presented which use only the 22 items which are shared between the two factor models, facilitating direct model comparisons. This analytic strategy facilitates the investigation of model fit in the various countries and the extent to which important model parameters remain consistent. Asparouhov and Muthén (2014) noted that conducting partial invariance analysis through multi-group CFA in studies with many groups can be impractical as following the modification indices in reaching measurement invariance can lead to the wrong model, due to scalar model being significantly different than the true model. For that reason, they suggested employing an alignment optimization procedure, which does not require exact measurement invariance to allow for estimation of group-specific factor means and variances.

All these stages consisted of analysing model fit and factor loadings using confirmatory factor analysis. Model fit was assessed using a battery of fit statistics. Given the lack of consensus on the appropriate use of fit statistics, the TLI, CFI, chi square, RMSEA, SRMR and AGFI were reported. It should be noted that reported fit indices are of the scaled/robust type, which are more robust to violations of normality than standard indices.

Despite being an inferential test, given the sensitivity to sample size

Table 1

Item configuration of original SD4 and Crawford et al. (2025) Dark Five model.

Code	Item	SD4 scale	D5 scale
M1	It's not wise to let people know your secrets.	Machiavellianism	Coalition building
M2	Whatever it takes, you must get the important people on your side.	Machiavellianism	Coalition building
M3	Avoid direct conflict with others because they may be useful in the future.	Machiavellianism	Coalition building
M4	Keep a low profile if you want to get your way.	Machiavellianism	Coalition building
M5	Manipulating the situation takes planning.	Machiavellianism	–
M6	Flattery is a good way to get people on your side.	Machiavellianism	–
M7	I love it when a tricky plan succeeds.	Machiavellianism	–
N1	People see me as a natural leader.	Narcissism	–
N2	I have a unique talent for persuading people.	Narcissism	–
N3	Group activities tend to be dull without me.	Narcissism	Grandiose exhibitionism
N4	I know that I am special because people keep telling me so.	Narcissism	Grandiose exhibitionism
N5	I have some exceptional qualities.	Narcissism	Grandiose exhibitionism
N6	I'm likely to become a future star in some area.	Narcissism	Grandiose exhibitionism
N7	I like to show off every now and then.	Narcissism	Grandiose exhibitionism
P1	People often say I'm out of control.	Psychopathy	Psychopathy
P2	I tend to fight against authorities and their rules.	Psychopathy	Psychopathy
P3	I've been in more fights than most people of my age and gender.	Psychopathy	Psychopathy
P4	I tend to dive in, then ask questions later.	Psychopathy	Psychopathy
P5	I've been in trouble with the law.	Psychopathy	Psychopathy
P6	I sometimes get into dangerous situations.	Psychopathy	Psychopathy
P7	People who mess with me always regret it.	Psychopathy	Psychopathy
S1	Watching a fist-fight excites me.	Sadism	Violent voyeurism
S2	I really enjoy violent films and video games.	Sadism	Violent voyeurism
S3	It's funny when idiots fall flat on their face.	Sadism	Indirect sadism
S4	I enjoy watching violent sports.	Sadism	Violent voyeurism
S5	Some people deserve to suffer.	Sadism	Indirect sadism
S6	Just for kicks, I've said mean things on social media.	Sadism	–
S7	I know how to hurt someone with words alone.	Sadism	Indirect sadism

(Hox & Bechger, 1998) and lack of a penalty for model complexity, chi-square was reported, however, it was interpreted cautiously, with more weight placed on the following parsimony adjusted indices, namely the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), the Tucker Lewis Index (TLI) and Adjusted Goodness of Fit index (AGFI). CFI is a measure of goodness of fit, relative to a simpler model. TLI follows similar rules to the CFI, however, places more emphasis on parsimony, relative to model fit, with similar cutoffs being adopted. The Standardised Root Mean Square Residual (SRMR) emphasises factor loadings to demonstrate how data does not fit the model and is more robust against violations of normality. RMSEA and its 95 % confidence intervals are also reported. Finally, the (AGFI) shows how closely the model comes to replicating the observed covariate matrix. Factor loadings were also investigated to provide extra information on

the model. Adopted cutoffs are provided in Table 2. Invariance between groups can be deemed to have occurred when CFI values change by >0.01 and RMSEA <0.015 (Chen, 2007; Cheung & Rensvold, 2002). Only samples that have at least one valid response in every category can be included in this analysis. Only samples that have at least one valid response in every category can be included in this analysis.

As per Bishop and Herron (2015) the assumption of equal distance between values was difficult to justify and as such the data was considered ordinal and WLSMV was adopted.

Due to the relative nature of factor loadings and the need to fix one value, the first significant item was selected to be fixed. In the Turkish and Romanian sample, item M1 did not exhibit statistically significant loadings on the respective factor. This can alter subsequent significance and factor loadings and as such M2 was selected to be fixed in these samples. This was done to protect the interpretability of the other factor loadings. This practice is common within these types of analyses as it does not influence the model fit indices and provides more meaningful information on loadings (Kline, 2023).

The analysis was performed in Jamovi software 2.3.28 for Windows 11 (The jamovi project, 2024), which is built on the *R* statistical language and employs the "lavaan" statistical package, and in R Studio (Posit Team, 2025).

2.4. Ethics

The Ethics Board of the Department of Psychology, University of Montenegro reviewed and approved this study (*reference 01-02/FF/25*). This study was preregistered: https://osf.io/4x2w5/.

3. Results

The fit statistics of the various samples are given in Table 3. Factor loadings for all samples are given in Supplementary material. All factor loadings were significant at the 0.001 alpha, except for M1 in Turkish and Romanian samples. In most samples (except the German, Dutch and Peruvian samples) M1 had the lowest factor loading.

As per the reasons given in the analysis section, the greatest emphasis is placed on CFI, TLI and AGFI as they represent a parsimony adjusted fit index. While investigating the non-grouped sample of all countries, the D5-SD4 exhibited acceptable fit indices when looking at CFI and SRMR, however, TLI cutoff was not met. In general, the D5-SD4 exhibited better fit than the 4FM, while the 4FM did not exhibit good fit. It was considered appropriate to not discount the 4FM at this stage as its performance in individual countries may have contributed to its poor fit.

The poorest fit was exhibited by the Peruvian sample, with both the 4FM and D5-SD4 exhibiting fit well below established cutoffs. The 4FM failed to exceed the >0.9 CFI cutoff in all samples except the Polish, UK and Romanian samples, and from this group, the British sample fell below this cutoff when using the more conservative TLI. The D5-SD4 performed better with good fit exhibited by the Serbian, Dutch, Turkish, Polish, British, Romanian and American samples. Of these, the Serbian, Dutch, Polish, British and Romanian samples exceeded the TLI

Table 2

Fit indices cut-offs.

Fit index	Conservative cutoff	Liberal cutoff
CFI &TLI	>0.95 (Hu & Bentler, 1999)	>0.9 (Iacobucci, 2010)
SRMR	<0.06 (Hu & Bentler, 1999)	<0.08 (Browne & Cudeck, 1993)
RMSEA	<0.06 (Hu & Bentler, 1999)	<0.08 (Browne & Cudeck, 1993)
AGFI	>0.9 (Hooper et al., 2008)	
Factor loadings	0.3–0.4 (minimum), >0.5 (Practically significant), >0.7 (gold standard) (Hair et al., 2019)	

Table 3

Fit indices for the 4FM and D5 in various worldwide samples.

Country	Model	$x2$	df	p	RMSEA	90 % CI		CFI	TLI	SRMR	AGFI
						Lower	Upper				
All countries	4FM	12,471	344	<0.001	0.070	0.069	0.071	0.857	0.843	0.063	0.963
	D5	6497	199	<0.001	0.067	0.065	0.068	0.903	0.888	0.054	0.974
Peru	4FM	1116	344	<0.001	0.060	0.056	0.064	0.765	0.742	0.073	0.934
	D5	620	199	<0.001	0.059	0.053	0.064	0.831	0.804	0.063	0.953
Serbia	4FM	1001	344	<0.001	0.063	0.058	0.067	0.864	0.850	0.077	0.955
	D5	512	199	<0.001	0.057	0.051	0.063	0.914	0.900	0.067	0.971
Canada	4FM	1126	344	<0.001	0.059	0.055	0.063	0.813	0.795	0.065	0.953
	D5	693	199	<0.001	0.058	0.053	0.063	0.855	0.831	0.057	0.968
Nether.	4FM	1416	344	<0.001	0.064	0.061	0.068	0.871	0.859	0.067	0.955
	D5	788	199	<0.001	0.063	0.058	0.067	0.916	0.902	0.059	0.968
Slovakia	4FM	3343	344	<0.001	0.068	0.066	0.070	0.832	0.815	0.071	0.952
	D5	1694	199	<0.001	0.063	0.060	0.066	0.888	0.870	0.058	0.967
Turkiye	4FM	952	344	<0.001	0.073	0.067	0.078	0.836	0.820	0.088	0.922
	D5	493	199	<0.001	0.066	0.059	0.074	0.903	0.887	0.075	0.947
Poland	4FM	673	344	<0.001	0.062	0.055	0.069	0.911	0.902	0.077	0.962
	D5	422	199	<0.001	0.067	0.058	0.076	0.923	0.910	0.074	0.967
Germany1	4FM	971	344	<0.001	0.055	0.051	0.059	0.774	0.752	0.066	0.947
	D5	526	199	<0.001	0.052	0.047	0.058	0.843	0.818	0.054	0.964
Germany2	4FM	1025	344	<0.001	0.063	0.059	0.067	0.805	0.786	0.076	0.942
	D5	579	199	<0.001	0.062	0.056	0.068	0.860	0.837	0.066	0.956
UK	4FM	668	344	<0.001	0.065	0.058	0.072	0.908	0.899	0.083	0.946
	D5	334	199	<0.001	0.055	0.045	0.065	0.956	0.949	0.066	0.968
Romania	4FM	647	344	<0.001	0.054	0.048	0.061	0.929	0.922	0.080	0.959
	D5	390	199	<0.001	0.057	0.048	0.065	0.943	0.934	0.073	0.968
USA	4FM	968	344	<0.001	0.065	0.060	0.070	0.861	0.848	0.079	0.942
	D5	563	199	<0.001	0.065	0.059	0.071	0.906	0.890	0.069	0.957

cutoff.

In relation to RMSEA, the D5-SD4 performed better than the 4FM in most instances, except Poland and Romania, and performed equally well in the USA. Fit indices indicated good fit (i.e. below 0.06) for the 4FM model in Canadian, one German and Romanian samples. However, these results were broadly consistent in the D5-SD4 with the addition of Serbian, Peruvian and British samples. When looking at SRMR, 4FM had acceptable values in all samples but Turkish, British and Romanian. Still, D5-SD4 had lower SRMR values in all samples, indicating better fit. Similar analysis using only the shared items of the two models are presented in supplementary material and show generally that the D5-SD4 exhibited better fit than the 4FM across all fit indices and that this represented a meaningful improvement to the model.

In general, the 4FM's poor performance was particularly notable in Peru and Germany, however, poor fit was demonstrated in several samples. It was deemed appropriate to bring forward the 4FM for configural invariance testing. Since certain response categories had 0 cases for one German and the Romanian sample, the configural invariance procedure couldn't be run as thresholds are necessary for conducting

invariance analysis with ordinal data (Vandenberg & Lance, 2000). The multi-group CFA was conducted on remaining 10 samples.

When determining configural invariance, as shown in Table 4 the D5-SD4 also exceeded the threshold for good fit, exceeding the CFI cutoff of above 0.9 showing that the underlying factor model remained appropriate for all groups. The 4FM did not exhibit good fit and as such configural invariance was not established and the model was not brought forward for more constrained models. The D5-SD4 complied with guidelines in Rutkowski and Svetina (2014) indicating that factor loadings were roughly equivalent between groups and as such metric invariance was established.

Testing for scalar invariance resulted in poorer fit and as the decrease in fit exceeded the guidelines a drop of 0.01 in CFI and 0.015 in RMSEA, scalar invariance was not demonstrated. To test if specific items were responsible for scalar invariance, partial invariance testing was conducted. Parameters were ranked according to their chi square adjustment and were sequentially freed. After releasing the top 25 % of parameters, CFI still exceeded the cutoffs required to demonstrate scalar invariance and as a result as per guidelines in Asparouhov and Muthén

Table 4

Fit statistics from invariance testing.

Model	$x2$	df	P	RMSEA	90 % CI Lower Upper	CFI	TLI	SRMR	AGFI
Configural Invariance (unconstrained model)									
4FM	13,489	3440	<0.001	0.068	0.0670 .069	0.858	0.844	0.079	0.924
5FM	7296	1980	<0.001	0.065	0.0640 .067	0.906	0.891	0.067	0.952
Metric Invariance (model with constrained loadings)									
5FM	7655	2142	<0.001	0.064	0.0650 .062	0.902	0.894	0.076	0.943
Scalar Invariance (model with constrained loadings and thresholds)									
5FM	16,879	2691	<0.001	0.091	0.0900 .093	0.750	0.786	0.072	0.902

(2014) the alignment method was used. Out of 880 thresholds and 220 loadings, 23.64 % of thresholds and 4.55 % of loadings (average 14.1 %) were found to be invariant. According to Muthén and Asparouhov (2014), if there are less than 25 % of non-invariant parameters, group means can be meaningfully compared across the groups, which this model attained.

4. Discussion

The purpose of this exploratory study was to examine if the so-called *Dark 5 Model of the SD4* (D5-SD4) provided a more appropriate dimensional representation of the SD4 than the original four factor model dark tetrad model across multiple populations and geographical locations. Our results showed that the D5-SD4 was consistently a better fit for the data than the traditional 4FM model on both the shared 22 items and the 28 items versions. The demonstration of metric invariance showed that both the underlying factor structure and the factor loadings of the D5-SD4 remained constant between groups, whereas similar invariance was not demonstrated by the 4FM. In the previous research (Dinić et al., 2024), where configural invariance was demonstrated for the 4FM, robust indices were not used, otherwise, the configural invariance would not be reached. The alignment procedure showed that less than 25 % of parameters were non-invariant, which is below the rule-of-thumb threshold (Muthén & Asparouhov, 2014), but it also showed that some items have low R^2, which represents non-invariance effect size estimates. Luong and Flake (2023) consider this to be a challenge for implementation of alignment procedure and encourage researchers to take a detailed approach when making decisions about non-invariability of items.

A key distinction between the 4FM and the D5-SD4 is how the sadism dimension is treated. In the D5-SD4, sadism is divided between the indirect sadism dimension and the violent voyeurism dimension, with Crawford et al. (2025) suggesting that the differentiation between these two concepts was the separation between those who enjoyed watching the infliction of pain on others compared to those who enjoyed *doing the inflicting* of pain. While these labels have yet to be verified empirically, they somewhat align with previously established distinction between direct and vicarious sadism (Paulhus & Jones, 2015; Dinić et al., 2020). Our findings support this separation as doing so improved model fit in all instances, regardless of the emphasis placed on parsimony by the various fit indices. This implies that there are structural issues with this factor, which should prompt caution from researchers and highlights the need for a reconceptualization of sadism. Moreover, it may be argued that like Crawford et al. (2025), our findings partially resolve the psycho-sadism jangle problem with the SD4 (Blötner et al., 2022). Future researchers may also consider further testing the sadism factors of the D5-SD4 alongside alternative measures of sadism (see Dinić et al., 2020) and simultaneously exploring predictive validity of relevant outcomes (e.g., aggression, cynicism).

There was an interesting finding regarding the Machiavellianism scale. Items loaded more strongly onto the D5-SD4 than the 4FM. However, the M1 item, which refers to 'keeping secrets', exhibited weak loadings in most samples and non-significant associations with coalition building in the Turkish and Romanian samples. The alignment procedure showed that out of four thresholds for this item, two were non-invariant for the Turkish sample. All five parameters, one loading and four thresholds, had very low non-invariance effect size estimates, labelled as R^2 by Asparouhov and Muthén (2014), which ranged from 0.000 to 0.294. Low R^2 values point to the fact that small amount of variability in the item parameter is explained by group mean differences. However, since there are no clear and strict guidelines for interpretation of R^2 values, this issue requires further investigation.

It is also important to recognise that within a conceptual framework of Machiavellianism, keeping secrets from others is an expected behaviour, however, with the amendment to this factor in the D5-SD4, conceptually this is now referred to as coalition building, whereby secret keeping may no longer be as salient.

4.1. Implications, limitations and future directions

These exploratory results suggest that the SD4 in its current form may not be as useful to researchers examining the Dark Tetrad traits. Finding that the D5-SD4 had better fit than the 4FM may be partially attributable to the removal of items as much as the different factor structure. While these findings should be interpreted with this consideration in mind similar results were identified in the supplementary analysis using only the shared items, somewhat mitigating this concern. We tentatively suggest further development of the scale based on presented findings about factor structure.

However, it is important to note some limitations. First, as previously stated, poor model fit is a well-recognised issue within dark traits research, however, the authors feel it important to recognise that model fit in itself is not the definitive arbiter of instrument validity. Revelle (2024) argued that the drive to improve factor structure may come at the expense of predictive validity, which is also an important (arguably more important) consideration. Unlike Crawford et al. (2025) there were no additional psychological measures to test criterion validity in the present study. Future researchers could assess the D5-SD4 in relation to key related factors (e.g., aggression, cyberbullying) to account for the contrast with the 4FM (e.g., violent voyeurism and indirect sadism vs sadism) to test this facet of instrument validity. Second, like Crawford et al. (2025), we noted some factors (e.g., coalition-building, indirect sadism) exhibiting low internal reliability, although, this is known to be a common concern when utilising brief measures (Ziegler et al., 2014). Finally, while a large range of countries were tested, some geographical areas were not represented (e.g., East Asia, Africa) or had small samples (e.g. Turkey, Poland, Romania). As such, researchers should continue to examine if the D5-SD4 factor structure remains appropriate in datasets from other populations. Finally, we recommend researchers to consider replacing the term "dark" with less stigmatizing and more precise terms such as "socially aversive" or "antagonistic".

5. Conclusion

Crawford et al. (2025) encouraged further investigation of the D5-SD4 structure in different samples. This exploratory paper provides preliminary evidence that the five-factor structure has a better fit than the original four-factor structure in 12 different samples from 11 countries worldwide. Further efforts should be directed at exploring covariates of indirect sadism and violent voyeurism, clarifying the differences among them for researchers and practitioners. Finally, alternative measures of the Dark Tetrad should be further explored to address both methodological and terminological of socially aversive traits.

CRediT authorship contribution statement

A. Dulović: Writing – original draft, Formal analysis, Data curation, Conceptualization. **C. Smyth:** Writing – original draft, Supervision, Methodology, Conceptualization. **D. Waldeck:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2025.113529.

Data availability

The data used in this paper is open source. All sources are referenced and it is explained in methods, that data is freely available in the repositories of the respective papers.

References

Aghababaei, N., & Błachnio, A. (2015). Well-being and the Dark Triad. *Personality and Individual Differences, 86*, 365–368. https://doi.org/10.1016/j.paid.2015.06.043

Asparouhov, T., & Muthén, B. (2014). Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495–508. https://doi.org/10.1080/10705511.2014.919210

Aytac, M. B. (2024). Investigating everyday sadistic consumption within Dark Tetrad. *Mendeley Data, V2*. https://doi.org/10.17632/9yc97z4tyv.2

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(1), 78–102. https://doi.org/10.1207/S15328007SEM0901_5

Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science, 8*(3), 297–302. https://doi.org/10.70252/LANZ1453

Blötner, C., & Grüning, D. J. (2023). An examination of the role of inverted Dark Tetrad items on structural properties and construct validity. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000805

Blötner, C., Webster, G. D., & Wongsomboon, V. (2023). Measurement invariance of the Short Dark Tetrad across cultures and genders. *European Journal of Psychological Assessment, 39*(5), 331–336. https://doi.org/10.1027/1015-5759/a000715. Advance online publication.

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

Branković, I., Dinić, B. M., & Jonason, P. K. (2023). How traditional stalking and cyberstalking correlate with the Dark Tetrad traits? *Current Psychology, 42*(30), 26238–26242. https://doi.org/10.1007/s12144-022-03681-z

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

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

Chester, D. S., Lynam, D. R., & Miller, J. D. (2025). It is past time to abandon the term "dark" as a descriptor of antagonistic traits. *Journal of Psychopathology and Clinical Science*, 1–9. https://doi.org/10.1037/abn0001024

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5

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

Crawford, J. J., Smyth, C., Crossey, B. P., & Waldeck, D. (2025). The Dark Five: A reconsideration of the Short Dark Tetrad (SD4). *Personality and Individual Differences, 235*. https://doi.org/10.1016/j.paid.2024.112971

Dinić, B. M., Buckels, E. E., & Kovačević, N. (2024). The Short Dark Tetrad (SD4): Measurement invariance across Serbian and Canadian samples and further validation. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000864

Dinić, B. M., Bulut Allred, T., Petrović, B., & Wertag, A. (2020). A test of three sadism measures: Short Sadistic Impulse Scale, varieties of sadistic tendencies, and assessment of sadistic personality. *Journal of Individual Differences, 41*(4), 219–227. https://doi.org/10.1027/1614-0001/a000319

Fino, E., Popuşoi, S. A., Holman, A. C., Iliceto, P., & Heym, N. (2025). Dimensionality, factorial invariance, and cross-cultural differential item functioning of the Short Dark Tetrad (SD4) in Italian, Romanian, and UK samples. *European Journal of Psychological Assessment, 41*(1), 13–23. https://doi.org/10.1027/1015-5759/a000775

Gajda, A., Moroń, M., Królik, M., Małuch, M., & Mraczek, M. (2022). The dark tetrad, cybervictimization, and cyberbullying: The role of moral disengagement. *Current Psychology*. https://doi.org/10.1007/s12144-022-03456-6

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate Data Analysis* (8th ed.). Pearson Prentice.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*, 53–60.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review, 11*, 354–373.

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

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*, 90–98.

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

Kay, C. S., & Arrow, H. (2022). Taking an elemental approach to the conceptualization and measurement of Machiavellianism, narcissism, and psychopathy. *Social and Personality Psychology Compass, 16*(4), Article e12662. https://doi.org/10.1111/spc3.12662

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Co.

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.

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*, Article 103455. https://doi.org/10.1016/j.actpsy.2021.103455

Luong, R., & Flake, J. K. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods, 28*(4), 905–924. https://doi.org/10.1037/met0000441

Međedović, J., Drndarević, N., & Ilijić, L. (2024). Unfavorable prison social climate links dark tetrad traits with self-reported institutional misconduct. *Personality and Individual Differences, 227*. https://doi.org/10.1016/j.paid.2024.112707

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology, 5*, 978. https://doi.org/10.3389/fpsyg.2014.00978

Papageorgiou, K. A., Denovan, A., & Dagnall, N. (2019). The positive effect of narcissism on depressive symptoms through mental toughness: Narcissism may be a dark trait but it does help with seeing the world less grey. *European Psychiatry, 55*, 74–79. https://doi.org/10.1016/j.eurpsy.2018.10.002

Paulhus, D. L., Buckels, E. E., Trapnell, P. D., & Jones, D. N. (2021). Screening for dark personalities: The Short Dark Tetrad (SD4). *European Journal of Psychological Assessment, 37*(3), 208–222. https://doi.org/10.1027/1015-5759/a000602

Paulhus, D. L., & Jones, D. N. (2015). Measures of dark personalities. In *Measures of personality and social psychological constructs* (pp. 562–594). Academic Press.

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

Posit Team. (2025). *RStudio: Integrated development environment for R*. PBC, Boston, MA: Posit Software (URL http://www.posit.co).

Rassin, E., de Roos, M., & van Dongen, J. (2024). Dark personality traits and deception, and the short dark tetrad (SD4) as integrity screening instrument. *Scientific Reports, 14*, 311. https://doi.org/10.1038/s41598-023-50968-7

Revelle, W. (2024). The seductive beauty of latent variable models: Or why I don't believe in the Easter Bunny. *Personality and Individual Differences, 221*, 1–17. https://doi.org/10.1016/j.paid.2024.112552

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*, 31–57. https://doi.org/10.1177/0013164413498257

van der Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Stanton, K., Gillikin, L., Willis, L., Woods-Gonzalez, R., Myntti, W., Paige, C., … Emery, N. N. (2025). Understanding broader community perspectives on the scientific accuracy and stigma of personality trait labels. *Personality Disorders: Theory, Research, and Treatment, 16*(1), 91–102.

Teličák, P., Halama, P., & Kohút, M. (2024). The relationship between dark tetrad and conspiracy beliefs. No consistent results across three different samples from Slovakia. *Current Psychology, 43*(11), 10448–10457. https://doi.org/10.1007/s12144-023-05171-2

The jamovi project. (2024). jamovi (Version 2.3.28) [Computer Software]. Retrieved from https://www.jamovi.org.

Thorndike, E. L. (1913). *An introduction to the theory of mental and social measurements* (2nd ed.). Teachers College Press. https://doi.org/10.1037/10866-000 (rev. and enl.).

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Waldeck, D., Berman-Roberts, E., Smyth, C., Riva, P., Adie, J., Holliman, A. J., & Tyndall, I. (2024). Unraveling perceived ostracism: The role of antagonistic traits and attachment orientation. *The Journal of Psychology*, 1–19. https://doi.org/10.1080/00223980.2024.2396837

Webster, G. D., & Wongsomboon, V. (2020, July 6). *The Hateful Eight (H8): An efficient multifaceted approach to the Short Dark Tetrad (SD4)*. https://doi.org/10.31234/osf.io/pr4u6

Zegarra-López, A. C., Uribe-Bravo, K. A., Berrocal-Aragonés, G., Prieto-Molinari, D. E., & López, M. C. (2024). Adaptation and psychometric properties of the Short Dark Tetrad in a Peruvian sample. *Testing, Psychometrics, Methodology in Applied Psychology, 31*, 171–187. https://doi.org/10.4473/TPM31.2.3

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them [Editorial]. *Journal of Individual Differences, 35*(4), 185–189. https://doi.org/10.1027/1614-0001/a000148